**Powerpoint 1**

?WorldPhones

Head(TootchGrowth)

Tail(ToothGrowth)

Dim(TootchGrowth) 🡪 aantal rijen en kolommen

TryCodeScool doorwerken

**Powerpoint 2**

Mean()

Median()

Sd(=standaarddeviatie )

Summary(ChickenWeight)

Axis Generic Function to Add an Axis to a Plot

abline Add Straight Lines to a Plot

arrows Add Arrows to a Plot

assocplot Association Plots

axTicks Compute Axis Tickmark Locations

axis Add an Axis to a Plot

axis.POSIXct Date and Date-time Plotting Functions

barplot Bar Plots

box Draw a Box around a Plot

boxplot Box Plots

boxplot.matrix Draw a Boxplot for each Column (Row) of a Matrix

bxp Draw Box Plots from Summaries

cdplot Conditional Density Plots

grid Add Grid to a Plot

hist Histograms

example(plot)

plot(database)

>plot(cars, xlab = "Speed (mph)", ylab = "Stopping distance (ft) ",sub=" bron:database in R")

title(main = "cars data from the year 1920")

>z <- lm(dist ~ speed, data = cars)

>abline(z)

**>plot(cars)**

**>abline(v=15)**

**>abline(h=40,col="red",lwd=3)**

**>abline(20,1)**

>cijfers<-c(7,7,8,9,6,4,6,7,8,5,4)

>barplot(cijfers)

>**resultaat<-c(7,5,8,9,6,4)**

**>names(resultaat)<-c("Marc","Piet","Maike","Jos","Klaas","Mieke")**

**>barplot(resultaat)**

> barplot(height=c(30,52,64,80),width=c(10,10,20,20),col=c(2,1,4,3))

> barplot(height=c(30,52,64,80),width=c(10,10,20,20),col=c(2,1))

**>binomialeverdeling<-dbinom(0:4,4,1/2)**

**>names(binomialeverdeling)<-c("0","1","2","3","4")**

**>barplot(binomialeverdeling,col="lavender",xlab="aantal successen")**

> barplot(cars$dist)

**> attach(cars)**

**> hist(dist)**

**Powerpoint 3**

**> boxplot(count ~ spray, data = InsectSprays)**

> boxplot(len~dose, data=ToothGrowth, boxwex =0.25, subset = supp == "VC", col = "yellow", main= "Guinea Pigs Tooth Growth", xlab = "VitaminC dose mg", ylab = "tooth length", xlim = c(0.5,3.5),ylim =c(0,35),yaxs ="i" )

> boxplot(len~dose, data = ToothGrowth, add =TRUE, boxwex = 0.25, subset = supp == 'OJ', col = "red")

> legend(2,9, c("Ascorbic acid","Orange juice"),fill = c("yellow","red"))

**> boxplot(len ~ dose, data = ToothGrowth, boxwex = 0.25, at = 1:3 - 0.2,**

**subset = supp == "VC", col = "yellow",**

**main = "Guinea Pigs' Tooth Growth", xlab = "Vitamin C dose mg",ylab = "tooth length",**

**xlim = c(0.5, 3.5), ylim = c(0, 35), yaxs = "i")**

> boxplot(len ~ dose, data = ToothGrowth, subset = supp == "VC", col = "yellow", main = "Guinea Pigs' Tooth Growth",xlab = "Vitamin C dose mg",ylab = "tooth length")

> boxplot(len ~ dose, data = ToothGrowth, add = TRUE, subset = supp == "OJ", col = "orange")

**Of**

> boxplot(len ~ dose, data = ToothGrowth, add = TRUE, boxwex = 0.25, at = 1:3 + 0.2, subset = supp == "OJ", col = "orange")

**> boxplot(weight~Diet, data=ChickWeight, subset = Time ==21, col = "yellow", main="Gewicht van kuikens in gram op dag 21 bij verschillende diëten", xlab="dieet", ylab="gewicht in gram", sub="bron:Package datasets in R")k**

> ChickWeightDay21 <- subset(ChickWeight, Time==21)

> Diet1 <- subset(Chi, ckWeightDay21, Diet==1)

> Diet2 <- subset(ChickWeightDay21, Diet==2)

> Diet3 <- subset(ChickWeightDay21, Diet==3)

> Diet4 <- subset(ChickWeightDay21, Diet==4)

> boxplot(Diet1$weight,Diet2$weight,Diet3$weight, Diet4$weight)

> title(main='Gewicht van kuikens in gram op dag 21 bij verschillende diëten')

**PowerPoint 4**

Openen van bestanden:

Read.csv(“path”)

Read.csv(“path”, header=TRUE)

**Read.csv(file.choose(), header=TRUE)**

bloemblaadjes <- read.csv(file.choose(),header=TRUE)

summary(bloemblaadjes)

Wanneer waardes niet gescheiden zijn door komma’s:

**flusub1 <- read.table(file.choose(). Sep= “\t”, header=TRUE)**

Zoek uit hoe vaak de verschillende gastheren voor komen in deze file, flusub1.

**> flusub1 <- read.table(file.choose(),sep="\t", header=TRUE)**

**> summary(flu)**

Voeg flusub1 en flusub2 samen in één dataframe, wat je griepvirus noemt.

**> flusub2<- read.table(file.choose(),sep="\t",header=TRUE)**

**> summary(flusub2)**

**> griepvirus <- merge(x=flu, y=flusub2)**

**> summary(griepvirus)**

Datasets ophalen via internet:

**zoopl<-read.table(file=url("http://www.stat.sfu.ca/~thompson/stat403-50/data/zooplankton.txt"),header=T)**

> a <- c(1, 3, NA, 7, 9)

> sum(a)

>>> NA

The sum is considered "not available" by default because one of the vector's values was NA.

> sum(a, na.rm = TRUE)

>>> [1] 20

**PowerPoint 5**

Voorbeeld dbinom functie

**>dbinom(x=3, size=20,prob=1/6) of >dbinom(3,20,1/6)**

**>>> 0,2378866**

>plot(dbinom(0:4,4,1/2), pch=16, ylim=c(0,0.4), xlim=c(0,4),col='blue',xlab= 'index', lab='dbinom(0:4,4,1/2)')

Binomial test 1

682 is success, nullhypthesis is p=3/4

>**binom.test(682,682+243, p = 3/4 )** of >**binom.test(c(682, 243), p = 3/4)**

Binomial test 2

**Binom.test(107,1000, p=0,1, alternative=”less”)**

Binomial test 3

**Binom.test(107,1000,p=0,1, alternative = “less”, conf.level=0.99)**

Chisq.test**:** chisq.test(x, y = NULL, correct = TRUE, p = rep(1/length(x), length(x)), rescale.p = FALSE, simulate.p.value = FALSE, B = 2000)

Chisq.test example 1

**> X <- c(A=20, B=15, C= 25)**

**> chisq.test(X)**

Chsq.test example 2

**> x <-c(89,37,30,28,2)**  #geobserveerde waarden

**> p<-c(0.3,0.2,0.2,0.19,0.11)** #is samen 1

**Chisq.test(x,p=p)**  #hier staan de verwachte verhoudingen

Chi-squared test for given probabilities data: x

X-squared = 39.328, df = 4, p-value = 5.96e-08

Of

**> x <-c(89,37,30,28,2)**  #geobserveerde waarden

**> p<-c(0.3,0.2,0.2,0.19,0.11)**  #is samen 1

**Chisq.test(x,p=p, rescale.p = TRUE)**  #hier staan de verwachte verhoudingen

X-squared = 9.9901, df = 4, p-value = 0.04059

Library(help=’stats’)

***?binomial***

dbinom(x, size, prob, log =FALSE)

dbinom(x=3, size=20, prob=1/6) = 0,2378866

dbinom(0:4,4,1/2)

sum(dbinom(0:4,4,1/2))

plot(dbinom(0:4,4,1/2),col="red")

> plot(dbinom(0:4,4,1/2),col="red", Xlim = 0:4)

pbinom(q, size, prob, lower.tail = TRUE, log.p=FALSE)

dbinom(3,20,1/6) = 0,2378866

qbinom(p,size, prob, lower.tail=TRUE, log.p=FALSE)

rbinom(n, size, prob)

***Binomial test example 1***

# Under (the assumption of) simple Mendelian inheritance, a cross between plants of two particular genotypes produces progeny 1/4 of which are "dwarf" and 3/4 of which are "giant“.

In an experiment to determine if this assumption is reasonable, a cross results in progeny

having 243 dwarf and 682 giant plants. If "giant" is taken as success, the null hypothesis is that

p = 3/4 and the alternative that p ≠ 3/4.

**binom.test(682,682+243, p = 3/4 )**

# or

**binom.test(c(682, 243), p = 3/4)**

# x number of successes, or a vector of length 2 giving the numbers of successes and failures, respectively. n number of trials; ignored if x has length 2.

**binom.test(c(682, 243), p = 3/4)** # Output:

Exact binomial test

data: c(682, 243)

number of successes = 682, number of trials = 925,

**p-value = 0.3825**

alternative hypothesis: true probability of success is not equal to 0.75

95 percent confidence interval: 0.7076683 0.7654066

sample estimates: probability of success 0.7372973

***?chisq.test***

chisq.test(x, y = NULL, correct = TRUE, p = rep(1/length(x), length(x)), + rescale.p = FALSE, imulate.p.value = FALSE, B = 2000)

**Arguments**

x a numeric vector or matrix. x and y can also both be factors.

y a numeric vector; ignored if x is a matrix. If x is a factor, y should be

a factor of the same length.

***Chisq.test example 1  
# goodness of fit***

**x <- c(A = 20, B = 15, C = 25)** # hier staan de geobserveerde waarden

**chisq.test(x)**

Chi-squared test for given probabilities

data: x X-squared = 2.5, df = 2, p-value = 0.2865

***Chisq.test example 2***

***#goodness of fit***

**x <- c(89,37,30,28,2)** # hier staan de geobserveerde waarden

**p <- c(0.30,0.20,0.20,0.19,0.11)** # is samen 1

# hier staan de verwachte verhoudingen

**chisq.test(x, p = p)**

Chi-squared test for given probabilities

data: x

X-squared = 39.328, df = 4, p-value = 5.96e-08

***Chisq.test example 3***

***#goodness of fit***

**x <- c(89,37,30,28,2)**

**p <- c(40,20,20,15,5)** # is samen 100 (%)

**chisq.test(x, p = p, rescale.p = TRUE)**

Chi-squared test for given probabilities

data: x

X-squared = 9.9901, df = 4, p-value = 0.04059

***Opdrachten bij de dataset airquality***

Informatie over de dataset airquality

Airquality{datasets}

A data frame with 154 observations on 6 variables:

1. Ozone numeric ozone (ppb)
2. Solar.R cumeric Solar R(lang)
3. Wind numberic Wind(mph)
4. Temp numeric Temperature (degrees F)
5. Month numberic Month (1-12)
6. Day numeric Day

Airquality$Wind

Attach(airquality) 🡪 namen van kolommen zal apart laten zien.

Attach(ariquality)

Cor.test(Wind.Ozone)

Of

Cor.test(airquality$Wind, airquality$Ozone)

Selecteer de subset, waarvoor de temperatuur onder de 70F en laat alleen rijnummer, Ozone en Temp zien.

**subset (airquality, Temp<70, select = c(Ozone, Temp))**

Selecteer de subset, waarvoor Wind>12,0 en laat alleen rijnummer, Ozone en Wind zien.

**Subset(airquality, Wind>12, select = c(Ozone, Wind))**

Selecteer de subset, waarvoor Day=1 en laat alleen rijnummer, Ozone en Temp en Day zien.

**Subset(airquality, Day=1, select = c(Ozone, Temp, Day**))

Selecteer de subset, waarvoor Day=1 en laat de kolom Temp weg.

**subset(airquality, Day==1, select= -Temp)**

Selecteer de kolommen Ozone tot en met Wind

**Subset(airquality, select = Ozone:Wind)**

Selecteer regel 8 tot en met 14 (week 2) uit het dataframe

**subset(airquality[8:14,])**

Selecteer de eerste tien zondagen uit het dataframe

**Subset(airquality[c(6,13,20,27,34,41,48,55,62,69)])**

**of**

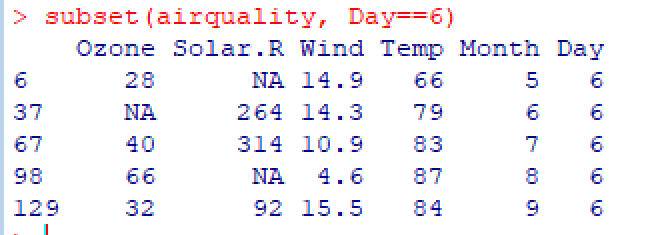
**Subset(airquality[seq(6,nrow(airquality),7),])**

Selecteer alle dagen uit het dataframe waarop Ozone de waarde Na heeft

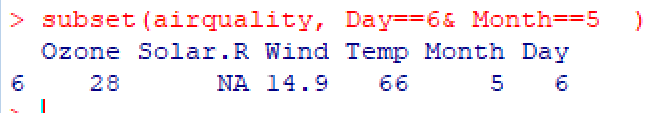
**Subset(airquality,is.na(Ozone))**

Laat het weer in NewYork zien in 19… op elke 6e dag van de maand.

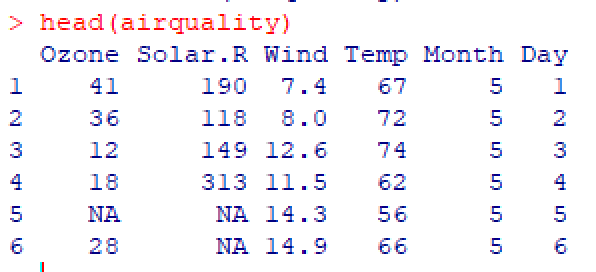
Subset(airquality, Day==6)



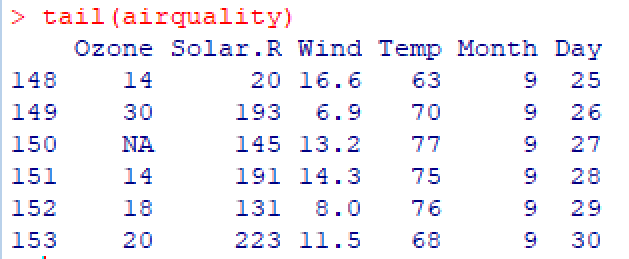
Laat het weer in NewYork zien in 19.. op de 6e dag van de 5e maand.



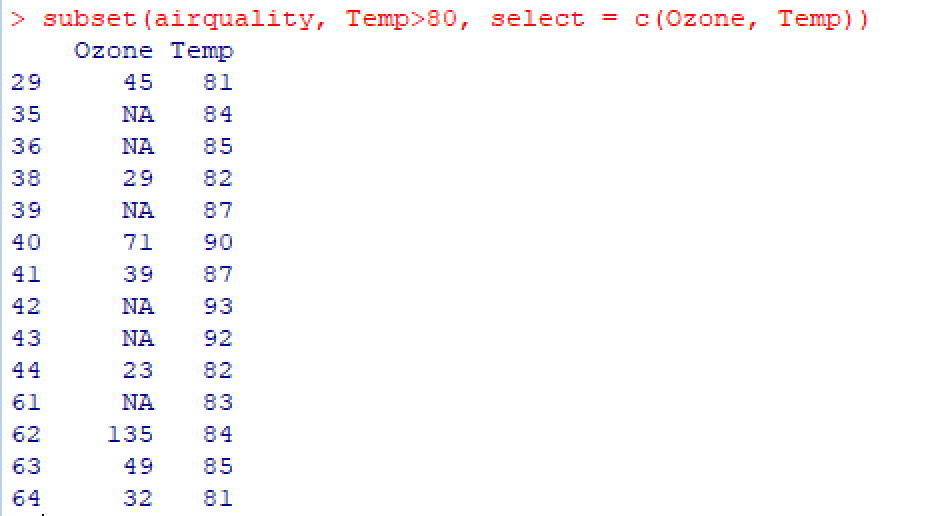
Head(airquality)



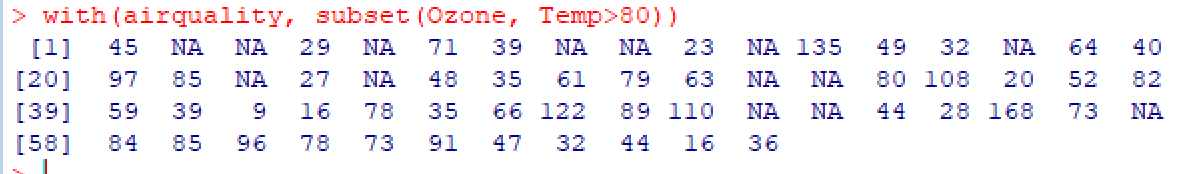
Tail(airquality)



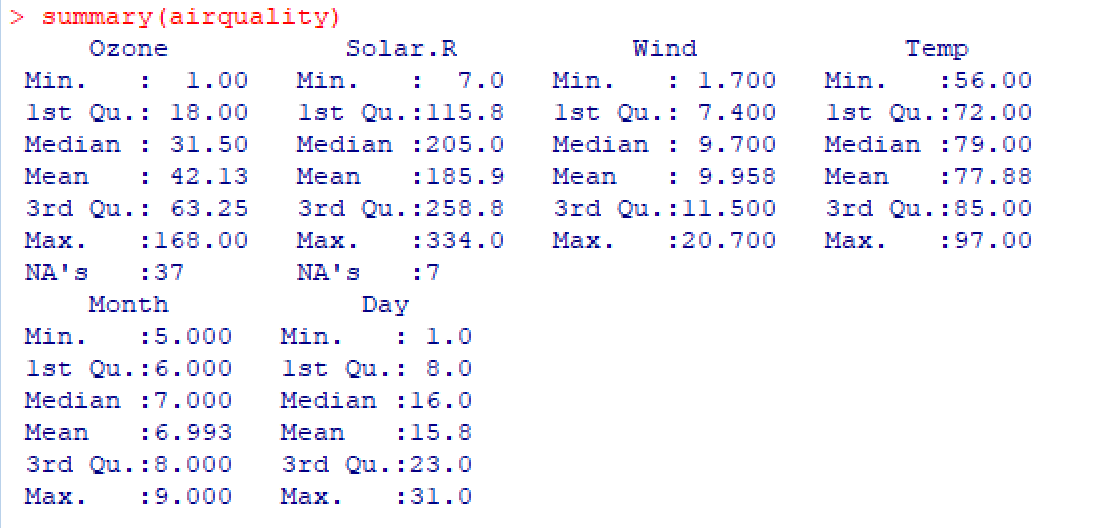
Subset(airquality, Temp>80, select = c(Ozone, Temp))



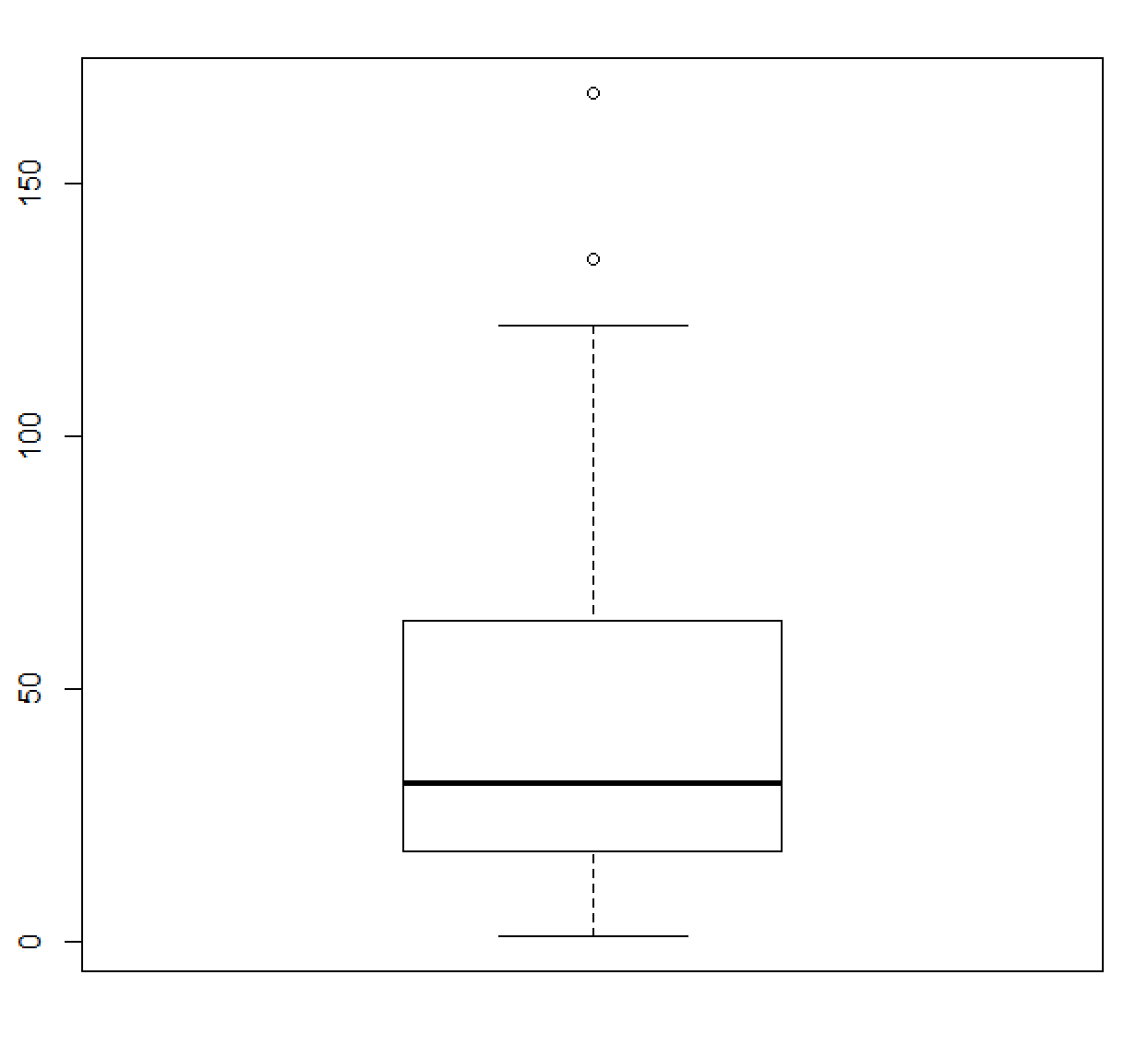
With(airquality, subset(Ozone, Temp>80))



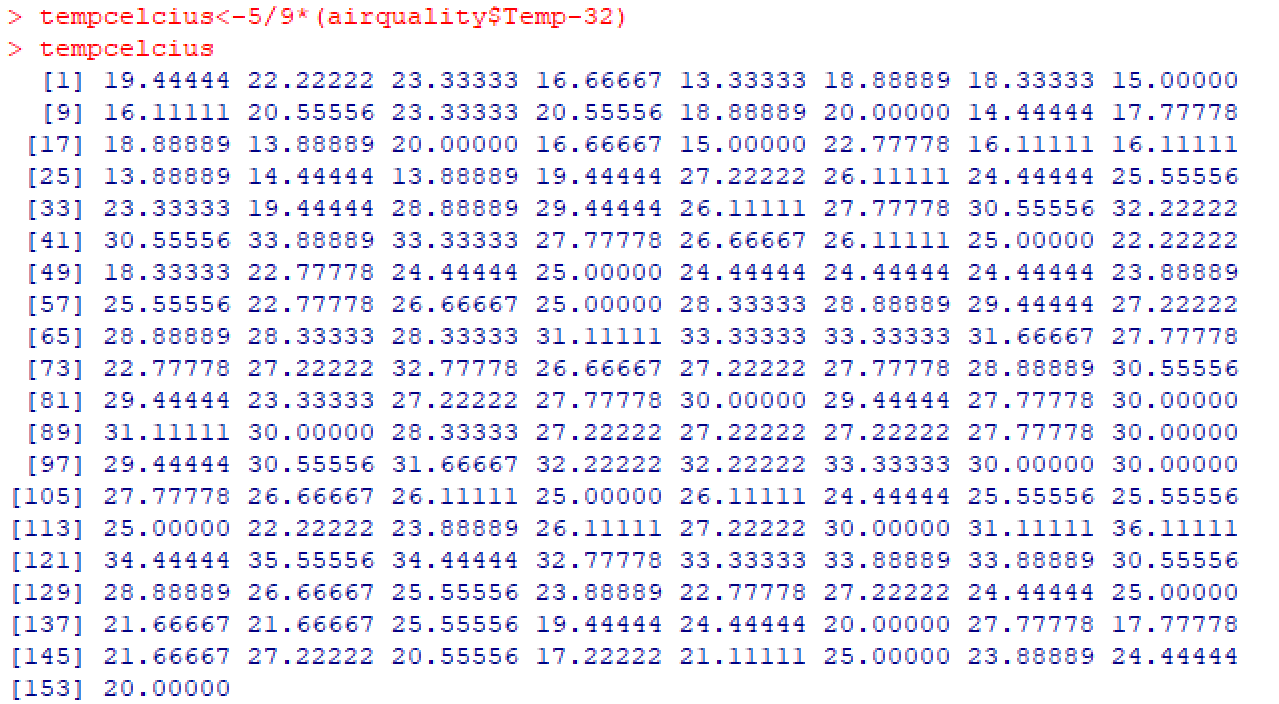
Summary(airquality)



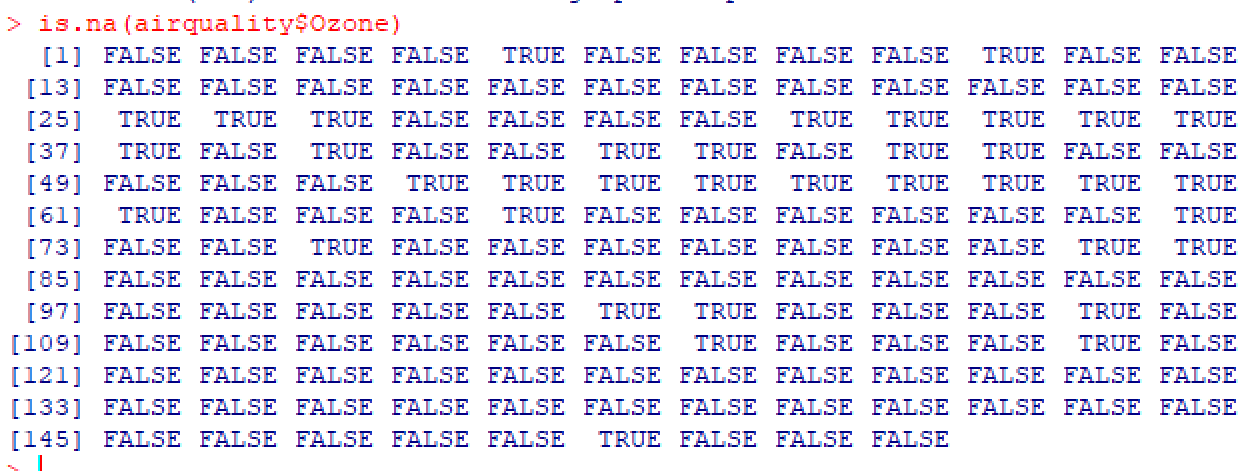
Boxplot(airquality$Ozone)



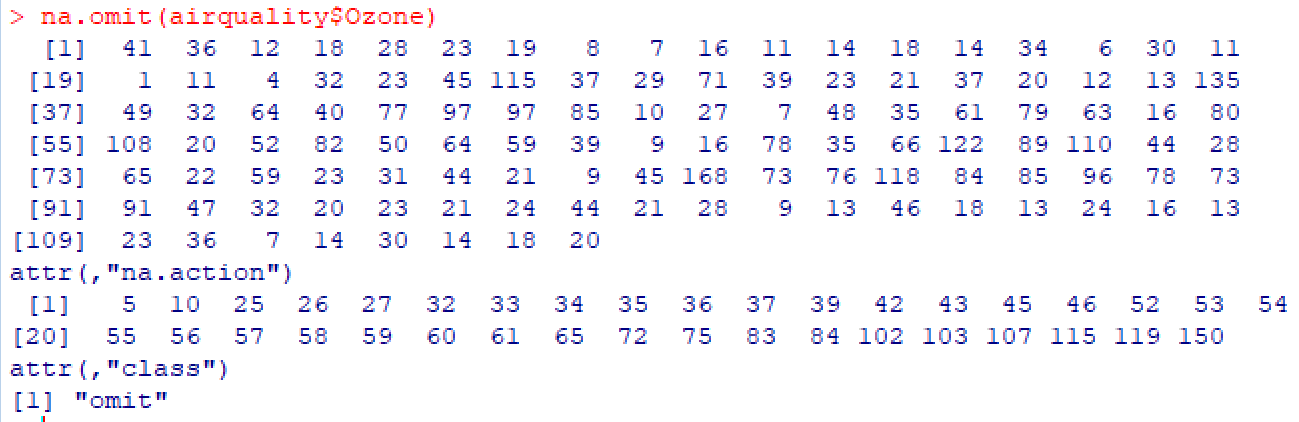
Tempcelcius<-5/9\*(airquality$Temp-32)



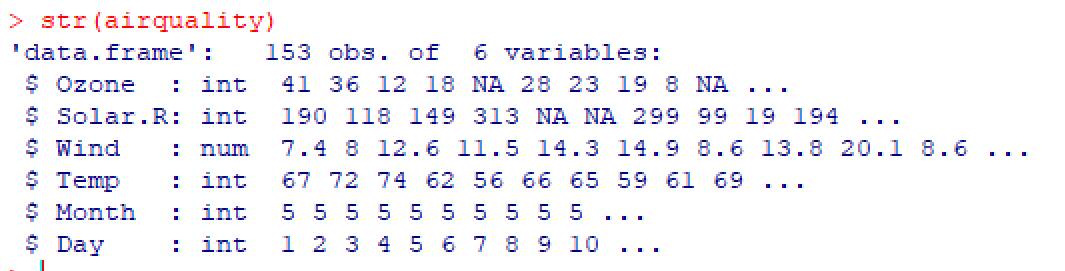
Is.na(airquality$Ozone)



Na.omit(airquality$Ozone)



Str(airquality)



**Powerpoint week 6**

**>cor.test(countries$GDP, countries$Piracy)**

Selecteer de subset, waarvoor de temperatuur onder de 70°F en laat alleen rijnummer, Ozone en Temp zien.

**>subset(airquality,Temp<70,select=c(Ozone,Temp))**

**Powerpoint week 7**

library(BiocInstaller) biocValid() ## R version 3.0 or later

> biocLite(c("geneplotter", "AnnotationDbi"))

**Cor.test(countries$GDP, countries$Piracy)**

Pearson’s product-moment correlation data:

Countries$GDP and countries$Piracy

T = -14.8371 df = 107 p-value < 2.2e-16

Alternative hypothesis: true correlation is not equel to 0

95 percent confidence interval: -0,8736179 -0,7475690

Sample estimates; cor -0.8203183

**Attach(long)**

**TABlongc=table(Smoke,Caesarean)**

**Chisq.est(TABlongc,cor**

**rect=T)**

> plot(Ozone~Wind, data = airquality, main = 'Ozon uitgezet tegen de Wind', sub = 'Airquality Database', cex = 1, cex.main = 2, cex.lab = 2)

**Attach(airquality)**

**Plot(Wind,Ozone, main=’Dagelijkse luchtkwaliteitmetingen in New York van 1 mei tot 30 seqtember 1973’, xlab=’windsnelheid in mph’, ylab= ‘ozon in ppb’, sub= ‘bron: package datasets in R’)**

plot(Wind,Ozone, main= 'Dagelijkse luchtwaliteitemetingen in NewUork van 1 mei tot 30 september 1973',xlab='Windsnelheid in mph', ylab='Ozon in ppb', sub='package datasets in R', col.main= 'darkblue', cex.main=3,cex.lab=2, abline(h=40,col='red', lwd=4), abline(v=15.1,col='darkgreen',lwd=2))

**> plot(Wind,Ozone, main= 'Dagelijkse luchtwaliteitemetingen in NewUork van 1 mei tot 30 september 1973',xlab='Windsnelheid in mph', ylab='Ozon in ppb', sub='package datasets in R', col.main= 'darkblue', cex.main=3,cex.lab=2, abline(h=40,col='red', lwd=4), abline(v=15,col='',lwd=2))**

**Oefentoets voor R**

**Package(library=’MASS’)**

Opdracht 1 a

Bestand ophalen uit datasets,

**C<- InsectSprays[25:36,]**

**D<-InsectSprays[37:48,]**

**Summary(C)**

**Summary(D)**

Odpracht 1 b

**Boxplot(count~spray, data = InsectSprays)**

Opdracht 1 c

**boxplot(count~spray, data = InsectSprays, main = 'Grafiek InsectSprays', ylab = 'Aantal insecten', xlab ='Soort Spray', sub = 'bron: Biometrika, 28,243-262')**

opdracht 1 d

Spray CDE in een groep&sprayABF in een groep.

* De plotjes in elke groep overlappen elkaar
* Interkwatielafstand is in de groep van sprayABF vele groter dan de groep van sprays CDE

Opdracht 1 e

**CDE <- InsectSprays[25:60]**

**> boxplot(count~spray, data= CDE)**

Opdracht 1 f

Tekst langs de xas: aantal aangetroffen insecten

Tekst langs de yas: aantal percelen

Titel: histogram van aantal percelen met een bepaald aantal insecten bij gebruik van insecticiden.

**hist(InsectSprays$count,xlab ='aantal aangetroffen insecten', ylab= 'aantal + percelen', main='Histogram van aantal percelen met een bepaald aantal insecten + bij begruik van insecticiden', sub='bron: Biometrika, 29,243')**

opdracht 2 a

**plot(body~brain, data=Animals)**

**> attach(Animals)**

**> plot(body~brain)**

**> logbody<-log(body)**

**> logbrain<-log(brain)**

**> plot(logbody,logbrain)**

2 b

**subAnimals<-subset(Animals,body<9000 )**

opdracht 3 a

**>par(bg=’lightcyan')**

**> plot(Luggage.room~Rear.seat.room, data = Cars93)**

Of

**Attach(Cars93)**

**Par(bf=’lightcyan’)**

**Plot(Rear.seat.room, Luggage.room)**

Opdracht 3 b

**>attach(Cars93)**

**>Par(bf=’lightcyan’)**

**> plot(Rear.seat.room, Luggage.room)**

**> abline(v=30,col="red",lwd=5)**

Opdracht 3 c

**Deelsetc<-subset(Cars93,select=c(Type,Length,Luggage.room,Rear.seat.room,Luggage.room,Makelwd))**

**> set1<-subset(WorldPhones, select=S.Amer:Mid.Amer)**

**> boxplot(set1, boxwex= 0.5, col='lightblue', main='The number of telephones in various regions of the world (in thousands)', cex.main=1, col.main='darkblue', ylab = 'Number of phones', xlab = 'Regions', cex.lab=1, col.lab='darkblue')**

**> legend(0.5,1000, c('regions uitgezet tegenover het aantal phones'), fill=c('lightblue'))**

library(datasets) iris\_dataset <- iris print(summary(iris\_dataset))

#meerdere dingen boxplotten in een boxplot

par(mfrow=c(1,1)) boxplot(len ~ dose, data = ToothGrowth, boxwex = 0.25, at = 1:3 - 0.2, subset = supp == "VC", col = "yellow", main = "Guinea Pigs' Tooth Growth", xlab = "Vitamin C dose mg", ylab = "tooth length", xlim = c(0.5, 3.5), ylim = c(0, 35), yaxs = "i") boxplot(len ~ dose, data = ToothGrowth, add = TRUE, boxwex = 0.25, at = 1:3 + 0.2, subset = supp == "OJ", col = "orange")

################################################ #####

Sepal Length and Sepal Width Boxplot ##### ##############/##################################

par(mfrow=c(2,2))

boxplot(Sepal.Length, data=iris\_dataset, subset = Species == "setosa", main="Sepal Length tegenover Sepal Width", xlab="Sepal length", ylab="Sepal width", sub="bron: iris dataset R", col='LightGrey')

boxplot(Sepal.Width, data=iris\_dataset, subset = Species == "setosa", main="Petal Length tegenover Petal Width", xlab="Petal length", ylab="Petal width", sub="bron: iris dataset R", col='Lightblue')

boxplot(Petal.Length, data=iris\_dataset, subset = Species == "setosa", main="Petal Length tegenover Petal Width", xlab="Petal length", ylab="Petal width", sub="bron: iris dataset R", col='Lightblue')

boxplot(Petal.Width, data=iris\_dataset, subset = Species == "setosa", main="Petal Length tegenover Petal Width", xlab="Petal length", ylab="Petal width", sub="bron: iris dataset R", col='Lightblue')

############################# #####

3 soorten plotten

##### #############################

par(mfrow=c(1,1)) iris\_setosa = subset(iris\_dataset, iris\_datasetSpecies=="setosa")irisvirginica=subset(irisdataset,irisdatasetSpecies=="setosa") iris\_virginica = subset(iris\_dataset, iris\_dataset*Species*=="*setosa*")*iris*​*v*​​*irginica*=*subset*(*iris*​*d*​​*ataset*,*iris*​*d*​​*ataset*Species=="virginica") iris\_versicolor = subset(iris\_dataset, iris\_datasetSpecies=="versicolor")plot(irissetosaSpecies=="versicolor") plot(iris\_setosa*Species*=="*versicolor*")*plot*(*iris*​*s*​​*etosa*Petal.Width, iris\_setosaPetal.Length,col=′red′,pch=3,xlim=c(0,2.5),ylim=c(0,7),main="PetallengthtegenoverPetalWidthvoor3soorten",xlab="PetalWidth",ylab="PetalLength",sub="bron:IrisdatasetR")points(irisvirginicaPetal.Length, col='red', pch=3, xlim=c(0, 2.5), ylim=c(0, 7), main="Petal length tegenover Petal Width voor 3 soorten", xlab="Petal Width", ylab="Petal Length", sub="bron: Iris dataset R") points(iris\_virginica*Petal*.*Length*,*col*=​′​​*red*​′​​,*pch*=3,*xlim*=*c*(0,2.5),*ylim*=*c*(0,7),*main*="*PetallengthtegenoverPetalWidthvoor*3*soorten*",*xlab*="*PetalWidth*",*ylab*="*PetalLength*",*sub*="*bron*:*IrisdatasetR*")*points*(*iris*​*v*​​*irginica*Petal.Width, iris\_virginicaPetal.Length,col=′blue′,pch=19)points(irisversicolorPetal.Length, col='blue', pch=19) points(iris\_versicolor*Petal*.*Length*,*col*=​′​​*blue*​′​​,*pch*=19)*points*(*iris*​*v*​​*ersicolor*Petal.Width, iris\_versicolor$Petal.Length, col='green', pch=17)