ChickWeight\_Dataset.R

Innova

Tue Jan 17 11:32:28 2017

#ChickWeight Dataset  
library(ggplot2)  
library(reshape2)  
library(plotly)

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:plotly':  
##   
## select

library(caret)

## Loading required package: lattice

head(ChickWeight)

## Grouped Data: weight ~ Time | Chick  
## weight Time Chick Diet  
## 1 42 0 1 1  
## 2 51 2 1 1  
## 3 59 4 1 1  
## 4 64 6 1 1  
## 5 76 8 1 1  
## 6 93 10 1 1

ls.str(ChickWeight)

## Chick : Ord.factor w/ 50 levels "18"<"16"<"15"<..: 15 15 15 15 15 15 15 15 15 15 ...  
## Diet : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 1 ...  
## Time : num [1:578] 0 2 4 6 8 10 12 14 16 18 ...  
## weight : num [1:578] 42 51 59 64 76 93 106 125 149 171 ...

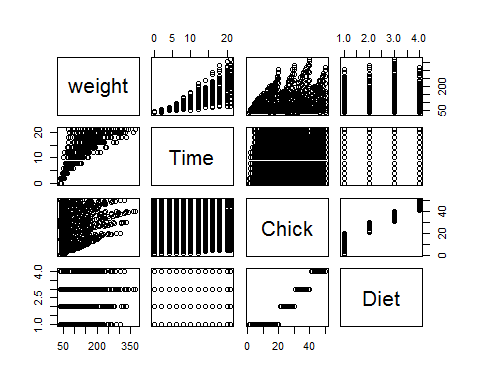
levels(ChickWeight$Diet)

## [1] "1" "2" "3" "4"

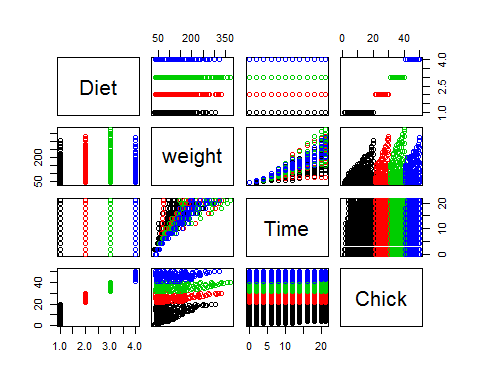
summary(ChickWeight)

## weight Time Chick Diet   
## Min. : 35.0 Min. : 0.00 13 : 12 1:220   
## 1st Qu.: 63.0 1st Qu.: 4.00 9 : 12 2:120   
## Median :103.0 Median :10.00 20 : 12 3:120   
## Mean :121.8 Mean :10.72 10 : 12 4:118   
## 3rd Qu.:163.8 3rd Qu.:16.00 17 : 12   
## Max. :373.0 Max. :21.00 19 : 12   
## (Other):506

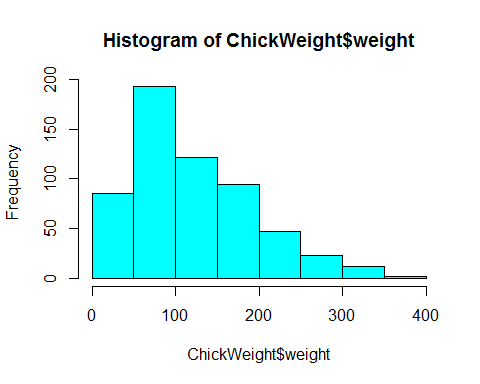
pairs(ChickWeight) #Scatterplot Matrix



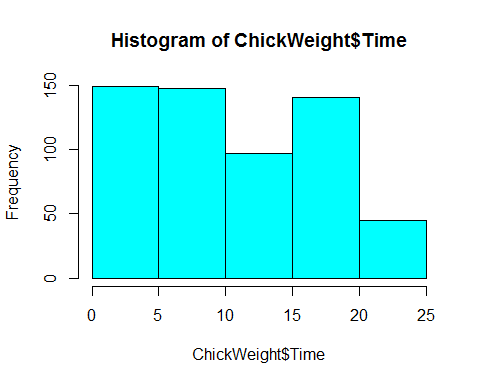
pairs(Diet~., data=ChickWeight, col=ChickWeight$Diet) #Scatterplot Matrix By Class



hist(ChickWeight$weight,col = "cyan")



hist(ChickWeight$Time,col = "cyan",breaks = 5)



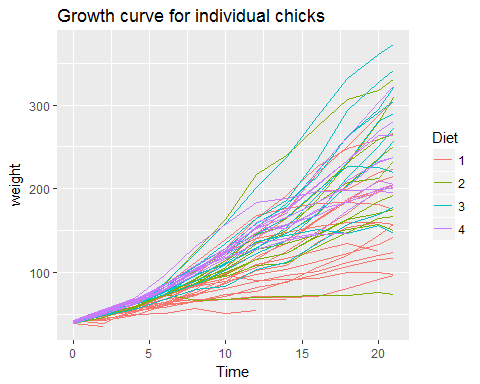
unique(ChickWeight$Chick)

## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23  
## [24] 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46  
## [47] 47 48 49 50  
## 50 Levels: 18 < 16 < 15 < 13 < 9 < 20 < 10 < 8 < 17 < 19 < 4 < ... < 48

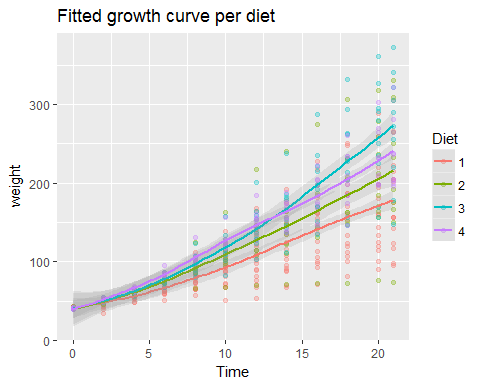
unique(ChickWeight$Diet)

## [1] 1 2 3 4  
## Levels: 1 2 3 4

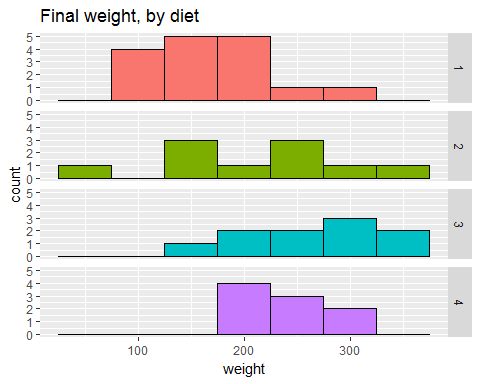
# Growth curve for individual chicks  
ggplot(ChickWeight, aes(x=Time, y=weight, colour=Diet, group=Chick)) +  
 geom\_line() +  
 ggtitle("Growth curve for individual chicks")



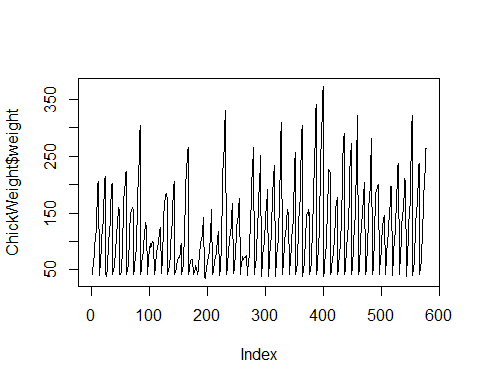
# Growth curve per diet  
ggplot(ChickWeight, aes(x=Time, y=weight, colour=Diet)) +  
 geom\_point(alpha=.3) +  
 geom\_smooth(alpha=.2, method="loess", size=1) +  
 ggtitle("Fitted growth curve per diet")



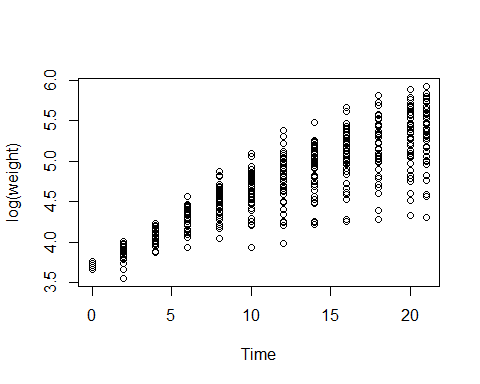
# Final weight, by diet  
ggplot(subset(ChickWeight, Time==21), aes(x=weight, fill=Diet)) +  
 geom\_histogram(colour="black", binwidth=50) +  
 facet\_grid(Diet ~ .) +  
 ggtitle("Final weight, by diet") +  
 theme(legend.position="none")



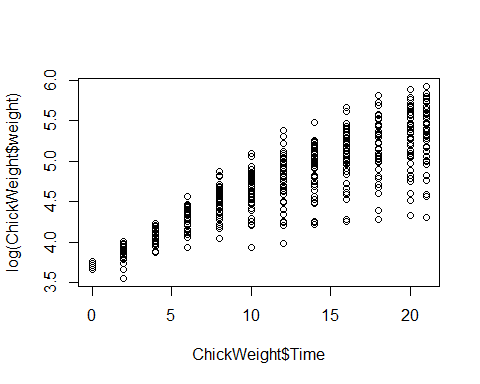
plot(ChickWeight$weight, type = "l")



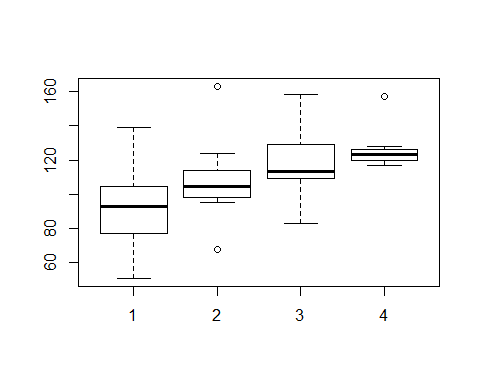
with(ChickWeight, plot(Time, log(weight)))



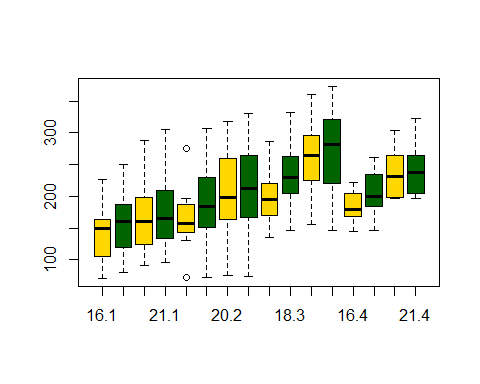
plot(ChickWeight$Time, log(ChickWeight$weight))



# informative graph of a selection of the data frame (time point 10)  
boxplot(weight ~ Diet,ChickWeight[ChickWeight$Time == 10,])



# take only part of the dataset  
temp = ChickWeight[ChickWeight$Time > 15,]  
# plot comparing weight vs 2 variables  
  
boxplot(weight ~ Time \* Diet,data=temp,col=c("gold","darkgreen"))



chick= ChickWeight$Chick  
weight=ChickWeight$weight  
  
plot(weight, chick )  
  
abline(lm( chick ~ weight)) #generate a linear regression model of the two variables

## Warning in model.response(mf, "numeric"): using type = "numeric" with a  
## factor response will be ignored

## Warning in Ops.ordered(y, z$residuals): '-' is not meaningful for ordered  
## factors

