InsectSprays\_Datset.R

Innova

Tue Jan 17 11:17:30 2017

#InsectSprays Datset  
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

data("InsectSprays")  
head(InsectSprays)

## count spray  
## 1 10 A  
## 2 7 A  
## 3 20 A  
## 4 14 A  
## 5 14 A  
## 6 12 A

ls.str(InsectSprays)

## count : num [1:72] 10 7 20 14 14 12 10 23 17 20 ...  
## spray : Factor w/ 6 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...

summary(InsectSprays)

## count spray   
## Min. : 0.00 A:12   
## 1st Qu.: 3.00 B:12   
## Median : 7.00 C:12   
## Mean : 9.50 D:12   
## 3rd Qu.:14.25 E:12   
## Max. :26.00 F:12

table(InsectSprays$spray)

##   
## A B C D E F   
## 12 12 12 12 12 12

insect=InsectSprays #assign the data values to an object called 'insect'  
library(data.table) #load in the necessary libraries

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:reshape2':  
##   
## dcast, melt

#find the mean, standard deviation, and standard error of each group (type of insect repellent) in the 'insect' data set  
insect\_table = data.table(insect)  
  
tapply(InsectSprays$count, InsectSprays$spray, mean)

## A B C D E F   
## 14.500000 15.333333 2.083333 4.916667 3.500000 16.666667

tapply(InsectSprays$count, InsectSprays$spray, quantile)

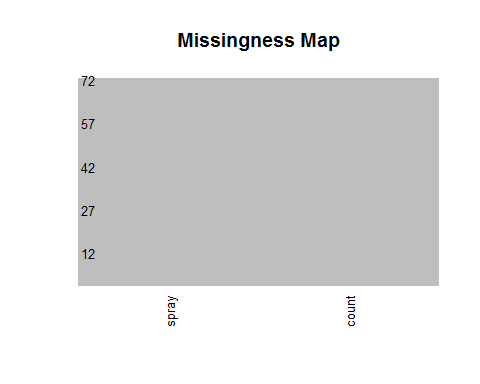
## $A  
## 0% 25% 50% 75% 100%   
## 7.00 11.50 14.00 17.75 23.00   
##   
## $B  
## 0% 25% 50% 75% 100%   
## 7.0 12.5 16.5 17.5 21.0   
##   
## $C  
## 0% 25% 50% 75% 100%   
## 0.0 1.0 1.5 3.0 7.0   
##   
## $D  
## 0% 25% 50% 75% 100%   
## 2.00 3.75 5.00 5.00 12.00   
##   
## $E  
## 0% 25% 50% 75% 100%   
## 1.00 2.75 3.00 5.00 6.00   
##   
## $F  
## 0% 25% 50% 75% 100%   
## 9.0 12.5 15.0 22.5 26.0

library(Amelia)

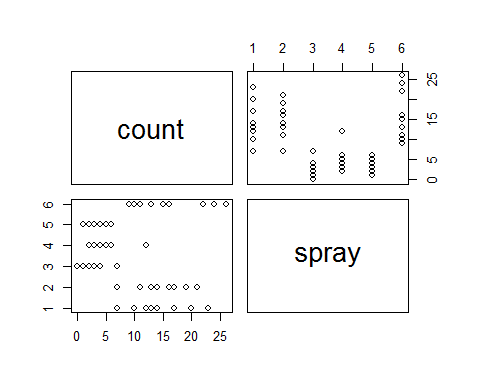
## Loading required package: Rcpp

## ##   
## ## Amelia II: Multiple Imputation  
## ## (Version 1.7.4, built: 2015-12-05)  
## ## Copyright (C) 2005-2017 James Honaker, Gary King and Matthew Blackwell  
## ## Refer to http://gking.harvard.edu/amelia/ for more information  
## ##

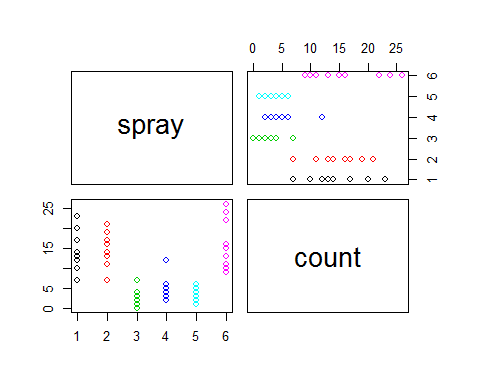
library(mlbench)  
missmap(InsectSprays, col=c("black", "grey"), legend=FALSE)



pairs(InsectSprays) ##Scatterplot Matrix



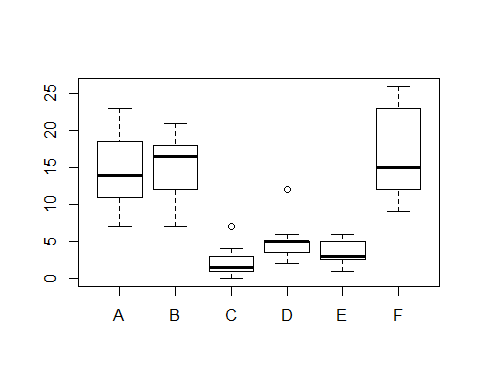
pairs(spray~., data = InsectSprays, col=InsectSprays$spray)



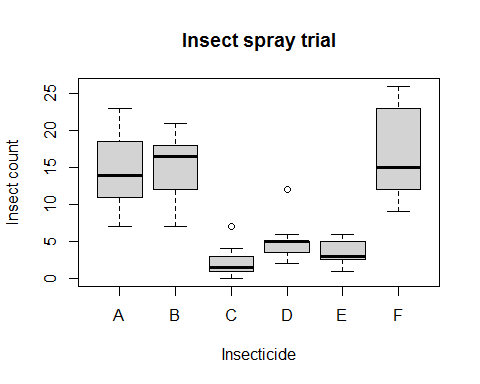
mean(InsectSprays$count, na.rm = TRUE)

## [1] 9.5

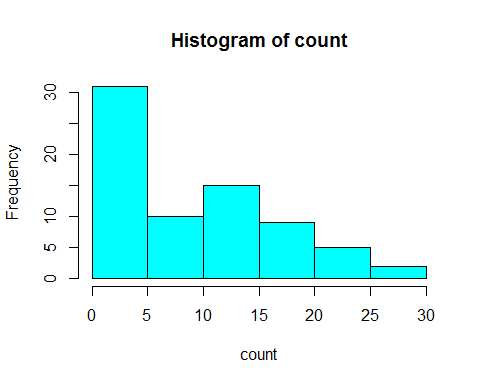
# plot  
boxplot(count ~ spray, data = InsectSprays)



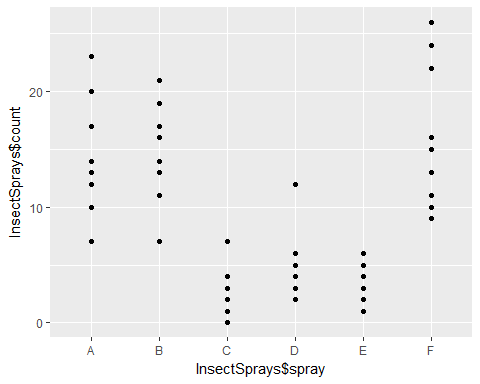
boxplot(count ~ spray, data = InsectSprays, col = "lightgray", main="Insect spray trial", xlab="Insecticide", ylab="Insect count")



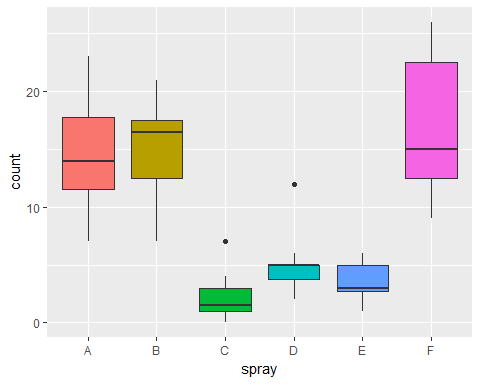
count = InsectSprays$count  
spray = InsectSprays$spray  
hist(count,col = "cyan")



qplot(InsectSprays$spray, InsectSprays$count)



#Create a boxplot to compare the different groups in dataset  
ggplot(InsectSprays,aes(x=spray,y=count,fill=spray))+geom\_boxplot()+guides(fill=FALSE)



model<-glm(count ~ spray - 1, data=InsectSprays, family = "poisson")  
model

##   
## Call: glm(formula = count ~ spray - 1, family = "poisson", data = InsectSprays)  
##   
## Coefficients:  
## sprayA sprayB sprayC sprayD sprayE sprayF   
## 2.674 2.730 0.734 1.593 1.253 2.813   
##   
## Degrees of Freedom: 72 Total (i.e. Null); 66 Residual  
## Null Deviance: 2265   
## Residual Deviance: 98.33 AIC: 376.6

summary(model)

##   
## Call:  
## glm(formula = count ~ spray - 1, family = "poisson", data = InsectSprays)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.3852 -0.8876 -0.1482 0.6063 2.6922   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## sprayA 2.67415 0.07581 35.274 < 2e-16 \*\*\*  
## sprayB 2.73003 0.07372 37.032 < 2e-16 \*\*\*  
## sprayC 0.73397 0.20000 3.670 0.000243 \*\*\*  
## sprayD 1.59263 0.13019 12.233 < 2e-16 \*\*\*  
## sprayE 1.25276 0.15430 8.119 4.71e-16 \*\*\*  
## sprayF 2.81341 0.07071 39.788 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 2264.808 on 72 degrees of freedom  
## Residual deviance: 98.329 on 66 degrees of freedom  
## AIC: 376.59  
##   
## Number of Fisher Scoring iterations: 5