Intro 2 R & Stats

Songpeng Zu

Tuesday, October 07, 2014

Intro 2 R

Intro 2 Stats.

Intro 2 R

Content

- R Basic
  1. Install R and Rstudio
  2. R classes
  3. R statements
  4. R functions and libraries
  5. **R I/O**
- R graph
  1. Basic graph
  2. ggplot
  3. **lattice**
- **Others**
  1. **Latex vs R markdown**
  2. **Rcpp**
  3. **R shiny**
  4. **R regular expression**

R Basic

- Install R
  `http://www.r-project.org/`
- Rstudio is a powerful IDE for R.
  `http://www.rstudio.com/`
- Working directory

```
getwd()
setwd("D:/")
```

R classes

character, numeric, integer, logical, **vector, matrix, factor, data frame, list.**

- vector

```
x <- c(10,3,2,1,NA); x[4]; mode(x); length(x)
```

- matrix

```
X <- matrix(1:12,nrow=3,byrow=TRUE,dimnames=NULL);X[1,]
## [1] 1 2 3 4
rbind(X,c(3,2,1,3))
##      [,1] [,2] [,3] [,4]
## [1,]    1    2    3    4
## [2,]    5    6    7    8
## [3,]    9   10   11   12
## [4,]    3    2    1    3
apply(X,2,mean)
## [1] 5 6 7 8
```

R classes

- factor: a compact way to handle categorical data.

```
sex <- c("M","F","M","F","F")
sex.factor <- as.factor(sex);table(sex.factor)
## sex.factor
## F M
## 3 2
height <- c(174,165,180,171,160)
tapply(height,sex.factor,mean)
##     F     M
## 165.3 177.0
gl(2,3)
## [1] 1 1 1 2 2 2
## Levels: 1 2
```

R classes

- list: a useful way to combine a collection of different objects.

```
family <- list(name="Fred",wife="Jane",
               children=c("XY","XX"))
family$name;family[[3]][2];family[["wife"]]
## [1] "Fred"
## [1] "XX"
## [1] "Jane"
family[1:2]
## $name
## [1] "Fred"
##
## $wife
## [1] "Jane"
unlist(family)
##       name       wife children1 children2
##     "Fred"     "Jane"       "XY"       "XX"
```

R classes

- data frame: a specific list of vectors and/or factors of the same length.

```
 df<-data.frame(
Name=c("Alice", "Becka", "James", "Jeffrey", "John"),
Sex=c("F", "F", "M", "M", "M"),
Age=c(13, 13, 12, 13, 12),
Height=c(56.5, 65.3, 57.3, 62.5, 59.0),
Weight=c(84.0, 98.0, 83.0, 84.0, 99.5)
); df
##      Name Sex Age Height Weight
## 1   Alice   F  13   56.5   84.0
## 2   Becka   F  13   65.3   98.0
## 3   James   M  12   57.3   83.0
## 4 Jeffrey   M  13   62.5   84.0
## 5    John   M  12   59.0   99.5
```

R statements

```
if(cond){expr}
else if(cond){expr}
else{expr}

for(var in seq){expr}

while(cond){expr}

break;next;
repeat {expr}
```

```
switch(3,1,2,3,4)
## [1] 3
switch("mean",mean=mean(c(1,3,2)),median=3)
## [1] 2
```

R functions and libraries

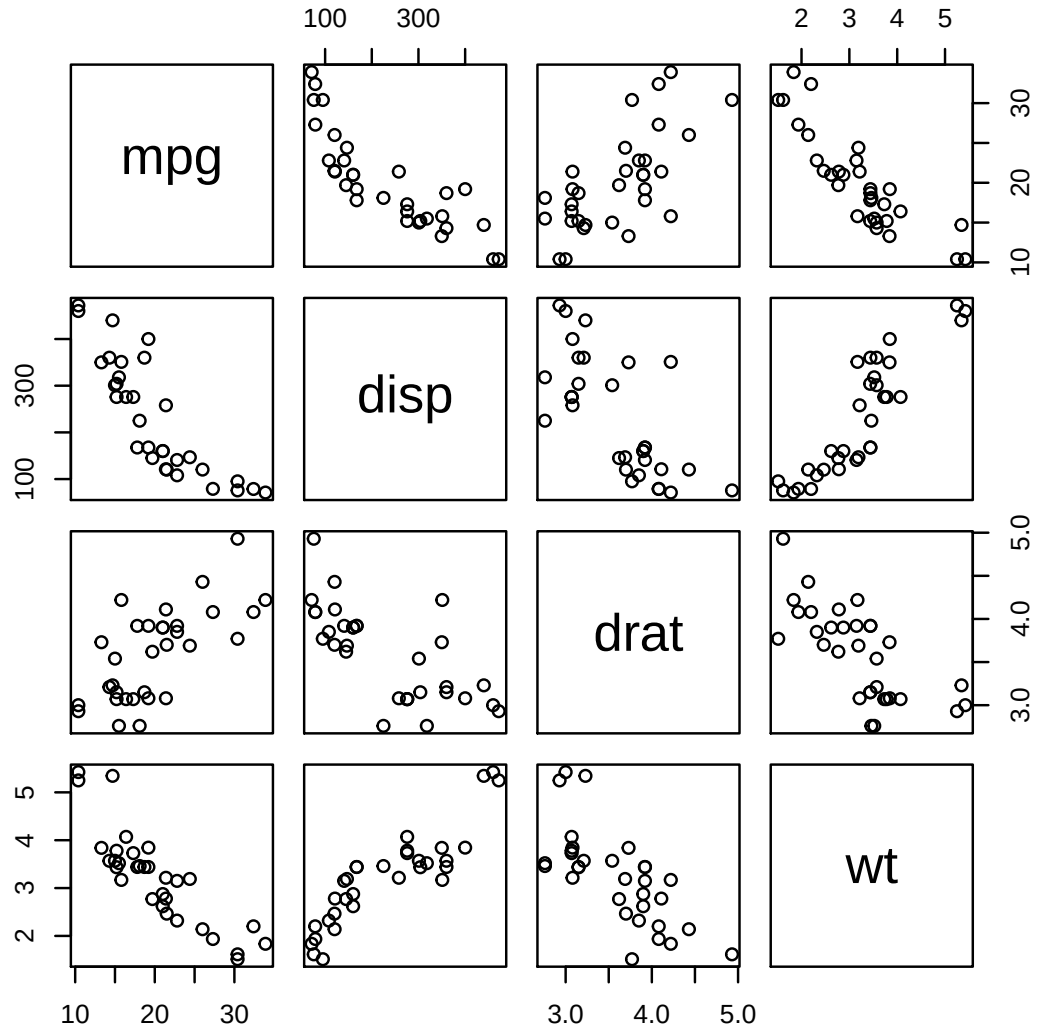
```
install.packages(packname)
library(packname)

funname <- function(arg1,arg2,...){
  statements
  return(objects)
}
```

```
centre <- function(x, type){
  switch(type,
         mean = mean(x),
         median = median(x))
}
x <- rcauchy(10)
centre(x, "mean")
## [1] -0.09672
```

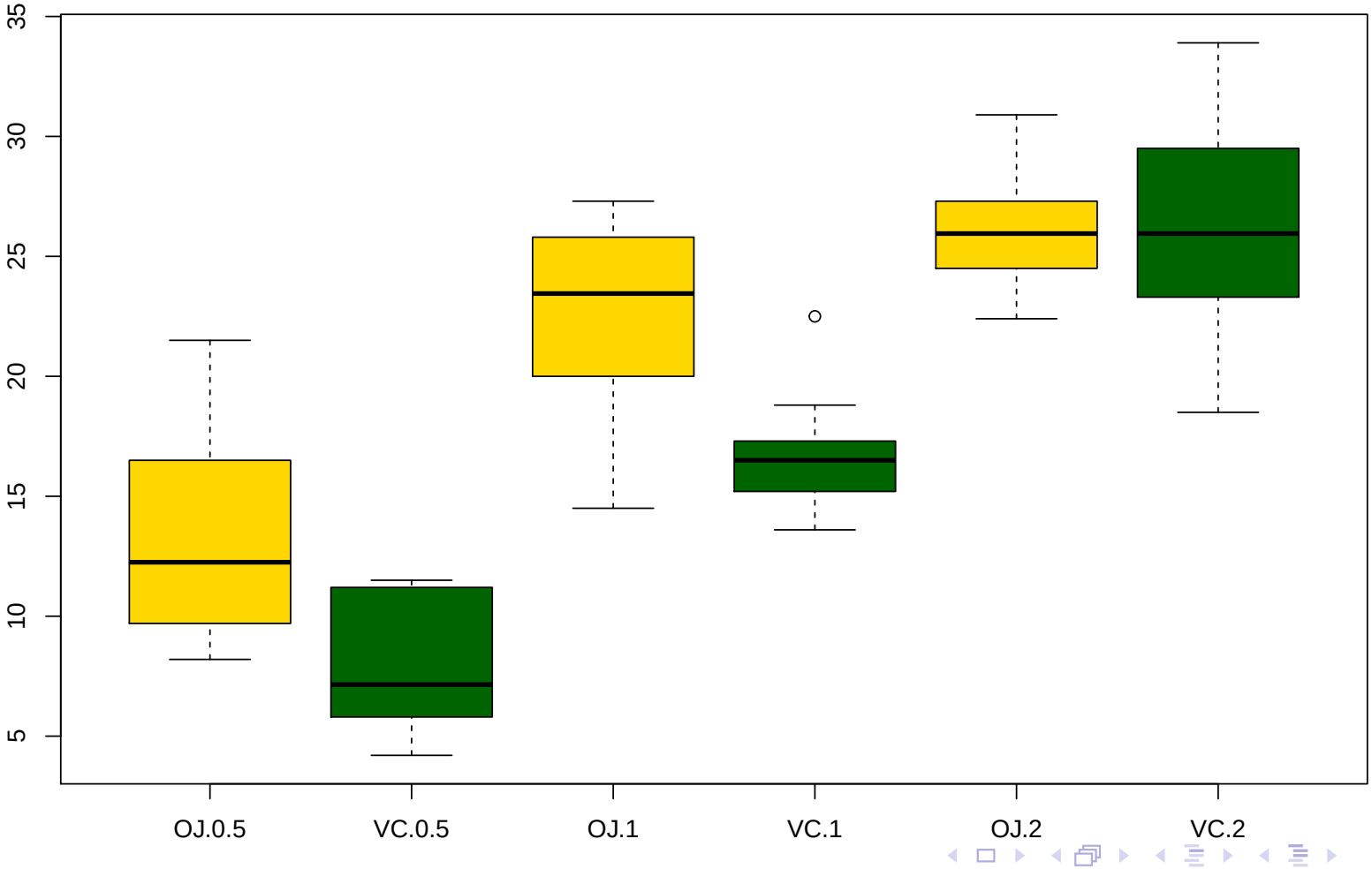
Basic graph

```
# Basic Scatterplot Matrix
pairs(~mpg+disp+drat+wt,data=mtcars)
```



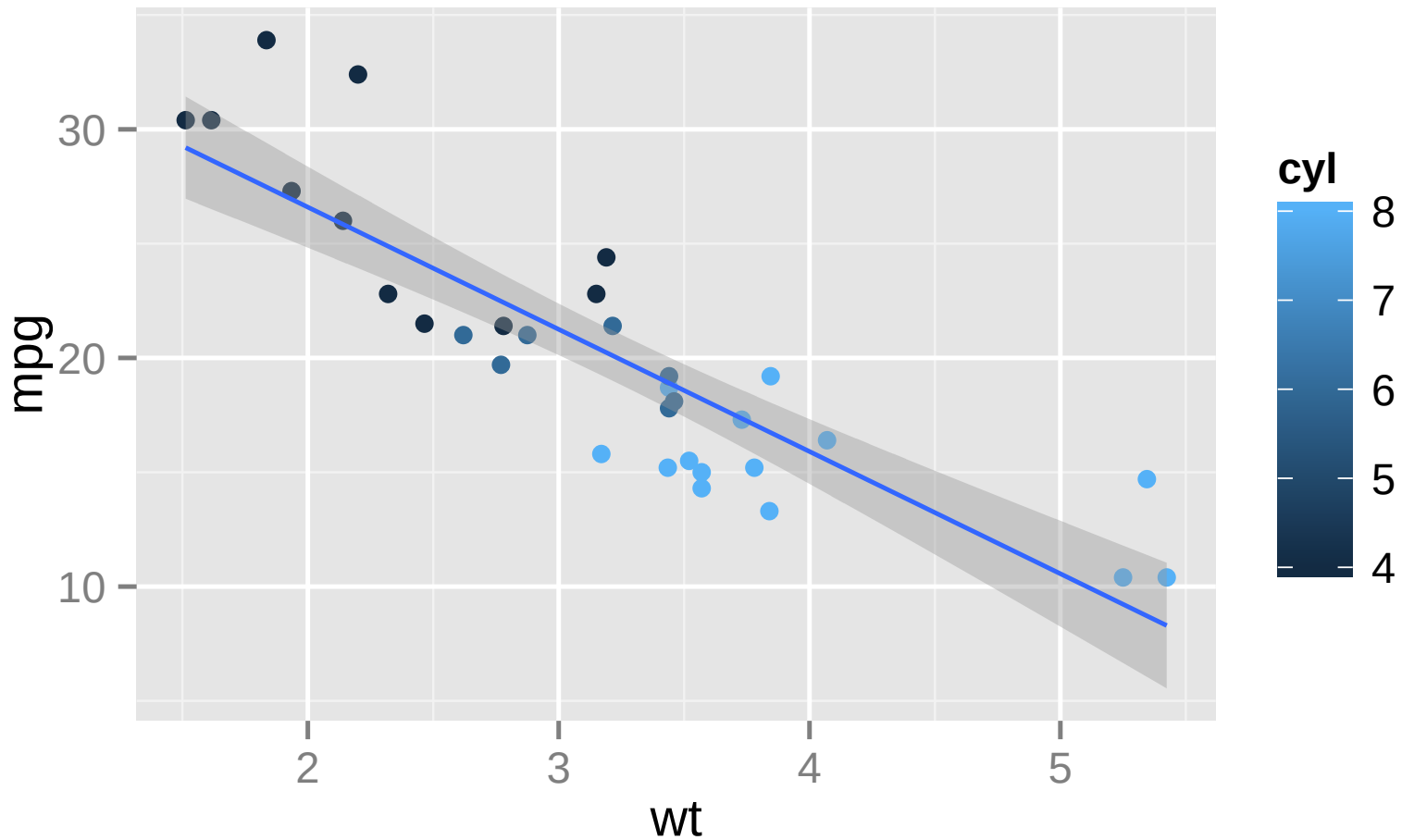
Basic graph

```
boxplot(len~supp*dose, data=ToothGrowth, notch=FALSE,
  col=(c("gold","darkgreen")))
```



ggplot

```
library(ggplot2)
qplot(wt, mpg, data=mtcars, geom=c("point", "smooth"),
   method="lm", formula=y~x, color=cyl)
```



Intro 2 Stats.

Content

- Bayesian Stats
  1.the Bayes Rule
  2.prior
  3.Hierachical Bayes
- Frequentist Stats
  1.Large sample theory for the MLE
  2.p-value, FDR

Bayesian Stats

- Bayes Rule

$$Pr(\theta|Data) \propto Pr(\theta)Pr(Data|\theta)$$

- prior

1. Conjugate prior distribution
2. Jeffreys prior

- Hierachical Bayes and an example.

Frequentist Stats

- Large sample theory for the MLE
- p-value
- FDR

Reference

- Peter Dalgaard, *Introductory Statistics with R*
- Winston Wang, *R Graphics Cookbook*
- `http://www.statmethods.net/`
- Kevin P.Murphy, *Machine Learning: A Probabilistic Perspective*