

R - Data visualisatie

J.J. van Nijnatten

Contents

Het doel van datavisualisatie	2
Wat wil je weergeven?	2
Hoe kies je de juiste manier van weergeven? (exp. design)	2
Essentiele onderdelen van datavisualisatie	2
gemiddelde	2
spreiding	2
legenda	2
titel	2
Introductie GGplot2 package	2
Installatie	2
Opbouw van figuren	2
Voorbeelden	2
Data-inspectie	2
Normaliteit	2
T-test	2
Correlatie	4
Regressie	4
One-way independent samples ANOVA	4
Factorial independent samples ANOVA	4
One-way repeated measures ANOVA	4
Factorial repeated measures ANOVA	5

contact: J.J.vanNijnatten@uva.nl

broncode: https://github.com/jonasvannijnatten/R_Data_Visualization

Het doel van datavisualisatie

Wat wil je weergeven?

Hoe kies je de juiste manier van weergeven? (exp. design)

Essentiele onderdelen van datavisualisatie

gemiddelde

spreiding

legenda

titel

Introductie GGplot2 package

Installatie

Het package downloaden & installeren:

```
install.packages(pkgs=c("ggplot2","Hmisc"), repos = "https://cloud.r-project.org")
```

Het package library activeren:

```
library(package="Hmisc")  
library(package="ggplot2")
```

Opbouw van figuren

Voorbeelden

Data-inspectie

Normaliteit

T-test

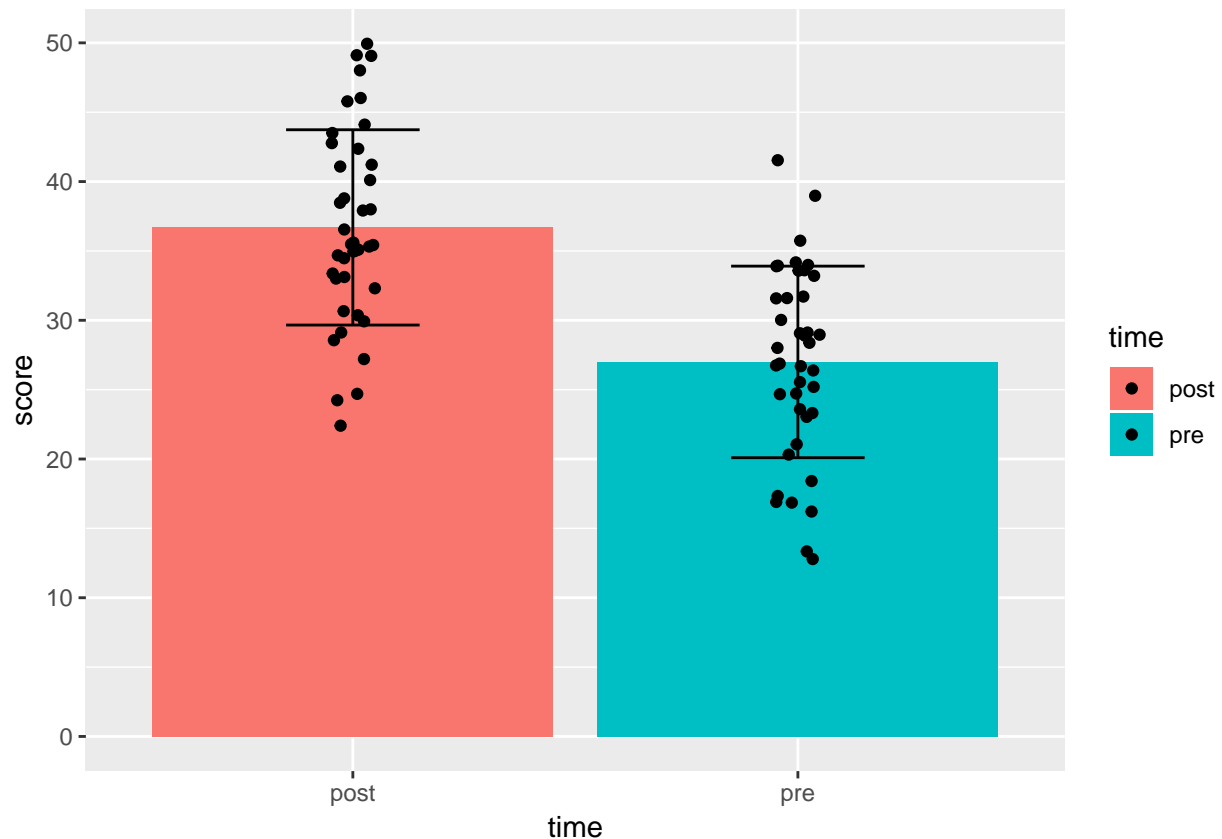
```
# generate data  
group1 = rnorm(n = 40, mean = 25, sd = 6.5)  
group2 = rnorm(n = 40, mean = 35, sd = 6.5)  
data.wide = data.frame(group1, group2)
```

```
# reshape data  
data.long = reshape(data = data.wide,  
                     direction = "long")
```

```
, varying = c("group1", "group2")
, v.names = "score"
, times = c('pre', 'post')
)
```

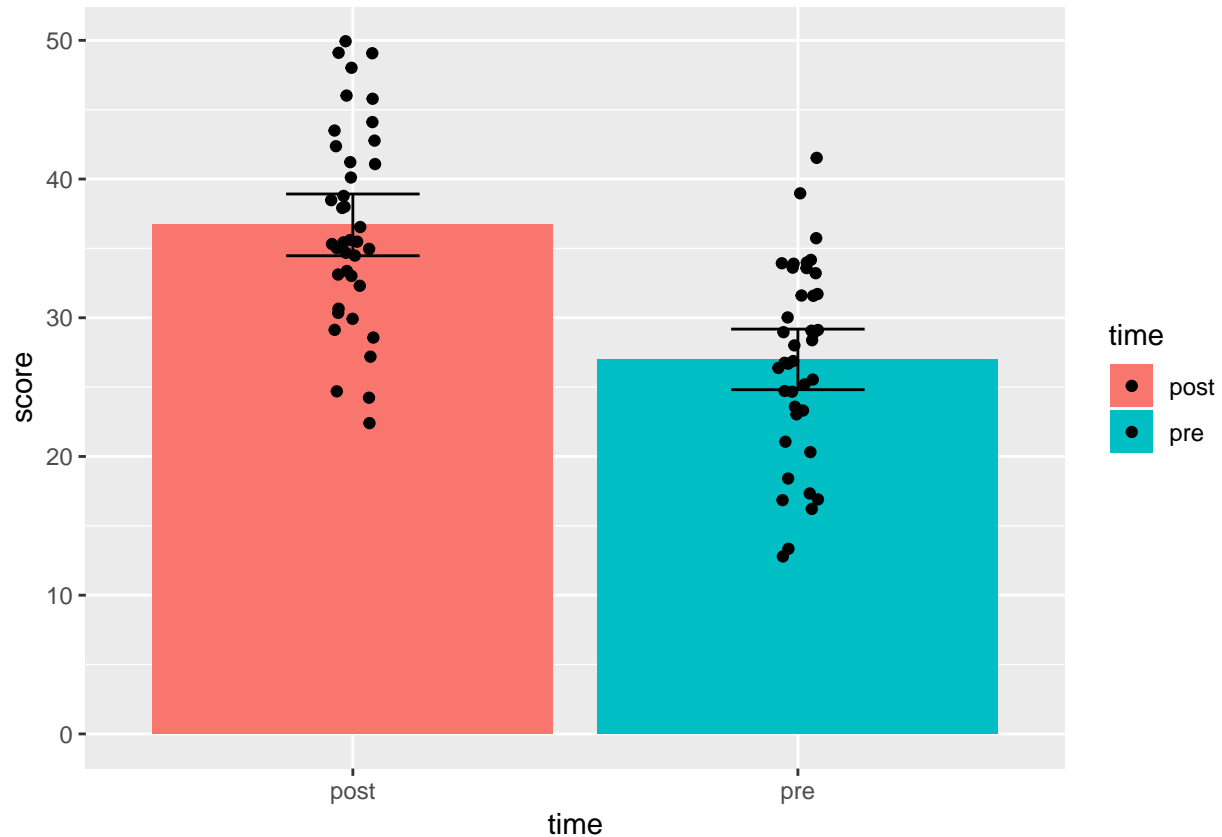
plot means and standard deviations

```
ggplot(data.long, aes(x=time, y=score, fill=time)) +
  geom_bar ( stat = "summary", fun.y = "mean" ) +
  geom_errorbar( stat = "summary", fun.data = "mean_sdl", fun.args = 1, width = 0.3 ) +
  geom_jitter ( width = .05 )
```



To plot standard errors instead of standard deviations replace “mean_sdl” with “mean_se”, and it is common use to plot 2 (or 1.96) times the standard error to get an 95% confidence interval, so replace “fun.arg = 1” with “fun.arg = 2”.

```
ggplot(data.long, aes(x=time, y=score, fill=time)) +
  geom_bar ( stat = "summary", fun.y = "mean" ) +
  geom_errorbar( stat = "summary", fun.data = "mean_se", fun.args = 2, width = 0.3 ) +
  geom_jitter ( width = .05 )
```



Correlatie

Regressie

One-way independent samples ANOVA

Factorial independent samples ANOVA

One-way repeated measures ANOVA

Generate dataset

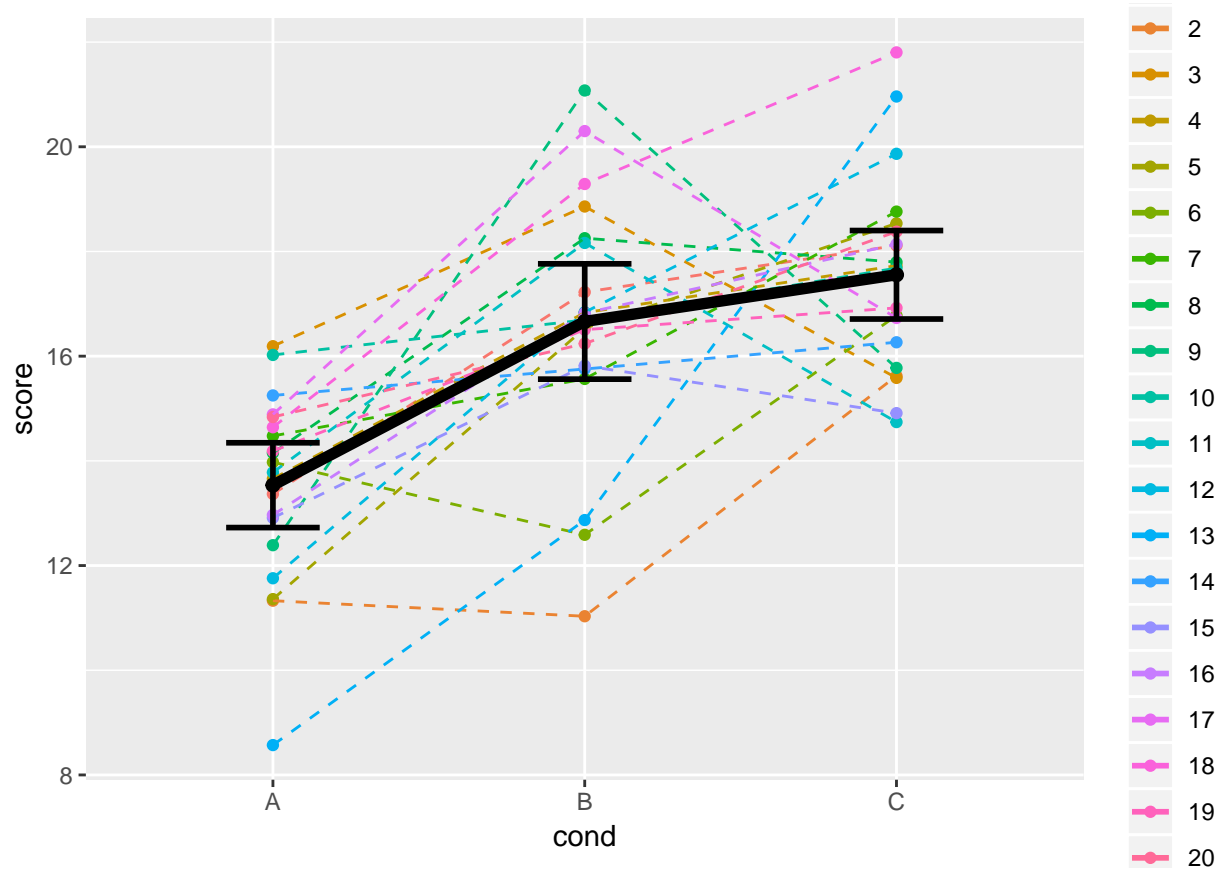
```
set.seed(01) # set seed
nrofsubs = 20 # set number of subjects
nrofconds = 3 # set number of conditions
subj = as.factor(rep(1:nrofsubs,nrofconds)) # create array with subject IDs
cond = as.factor(rep(LETTERS[1:nrofconds],each=nrofsubs)) # create array with condition values
score = as.vector( replicate(
  nrofconds , rnorm(n = nrofsubs, mean = sample(8,1)+10 , sd = sample(5,1) )
) ) # create array with measurement values
data.long = data.frame(subj, cond, score); # combine arrays into a data.frame
rm(list=setdiff(ls(), c("data.long", "nrofsubs","nrofconds"))) # delete arrays

ggplot(data.long, aes(x=cond, y=score, group=1, colour=subj)) +
  geom_point(aes(group=subj)) +
```

```

geom_line ( stat = "summary", fun.y = "mean", linetype= "dashed", aes(group=subj) ) +
geom_line ( stat = "summary", fun.y = "mean", colour = "black", linetype= "solid", size=2 ) +
geom_point ( stat = "summary", fun.y = "mean", colour = "black", size=2 ) +
geom_errorbar( stat = "summary", fun.data = "mean_se", fun.args = 2, width = 0.3, size=1 )

```



Factorial repeated measures ANOVA