Visualisation de données avec R

Marina Doucerain

marinadoucerain@gmail.com

- Rstudio
- Exploration & diagnostiques
- lattice
- ggplot2


RStudio

~/Documents/Workshops-training/R - Graphics - RStudio

File Edit Code View Plots Session Build Debug Tools Help

graphics.R* Untitled1*

Source on Save Run Source

```
31  SSW <- SS
32
33  write.table(ss, "graphics-data.txt")
34
35  summary(lm(salary.pre~visible+sex+edu+yrs*visible, data=ss))
36
37
38
39  ### Visual inspection of data ####
40
41  # Boxplots
42
43  boxplot(ss$salary.pre)
44  Boxplot(ss$salary.pre, id.n=Inf)
45  Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels
    =as.character(ss$visible))
46
47  # Correlation matrix
48
49  library(corrplot)
```

41:11 Visual inspection of data R Script

Console ~/Documents/Workshops-training/R - Graphics/

```
> Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels=as.cha
racter(ss$visible))
[1] "no" "no" "no" "no"
>
```

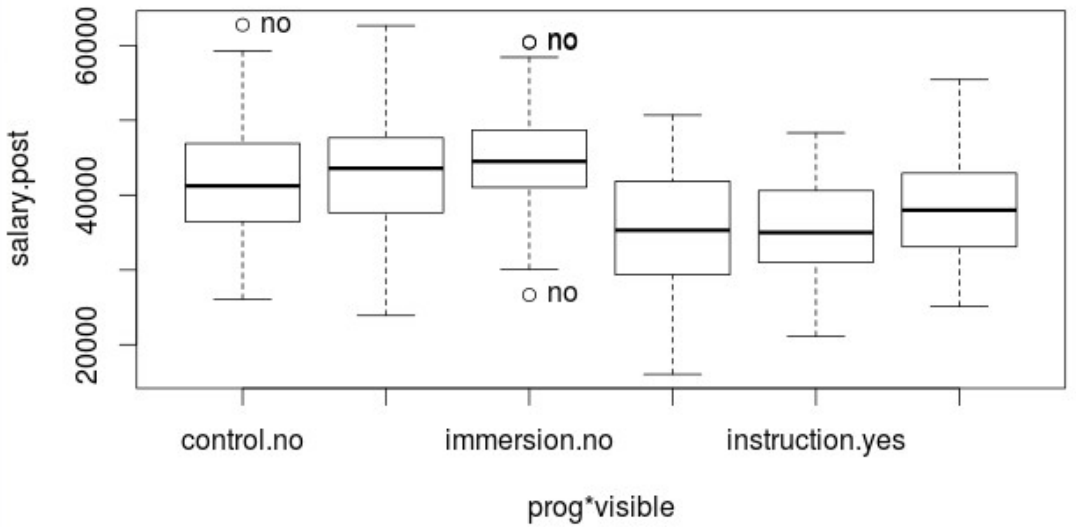
Environment History Help

R: HCL and HSV Color Palettes Find in Topic

```
## (in color and desaturated)
par(mar = rep(0, 4), mfrow = c(2, 2))
## rainbow color wheel
wheel(rainbow_hcl(12))
wheel(rainbow(12))
wheel(desaturate(rainbow_hcl(12)))
wheel(desaturate(rainbow(12)))
```

Files Plots Packages Viewer

Zoom Export Clear All



~/Documents/Workshops-training/R - Graphics - RStudio

File Edit Code View Plots Session Build Debug Tools Help

graphics.R* Untitled1*

Source on Save Run Source

```
ssw <- ss

write.table(ss, "graphics-data.txt")

summary(lm(salary.pre~visible+sex+edu+yrs*visible, data=ss))



### Visual inspection of data ####

# Boxplots

boxplot(ss$salary.pre)
Boxplot(ss$salary.pre, id.n=Inf)
Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels
=as.character(ss$visible))

# Correlation matrix

library(corrplot)
```

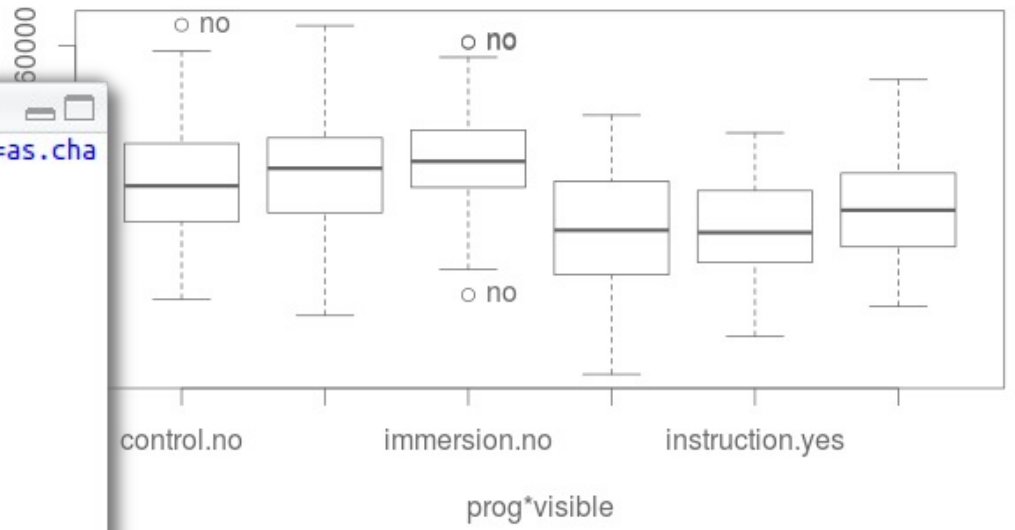
Environment History Help

R: HCL and HSV Color Palettes

```
## (in color and desaturated)
par(mar = rep(0, 4), mfrow = c(2, 2))
## rainbow color wheel
wheel(rainbow_hcl(12))
wheel(rainbow(12))
wheel(desaturate(rainbow_hcl(12)))
wheel(desaturate(rainbow(12)))
```

Files Plots Packages Viewer

Zoom Export Clear All

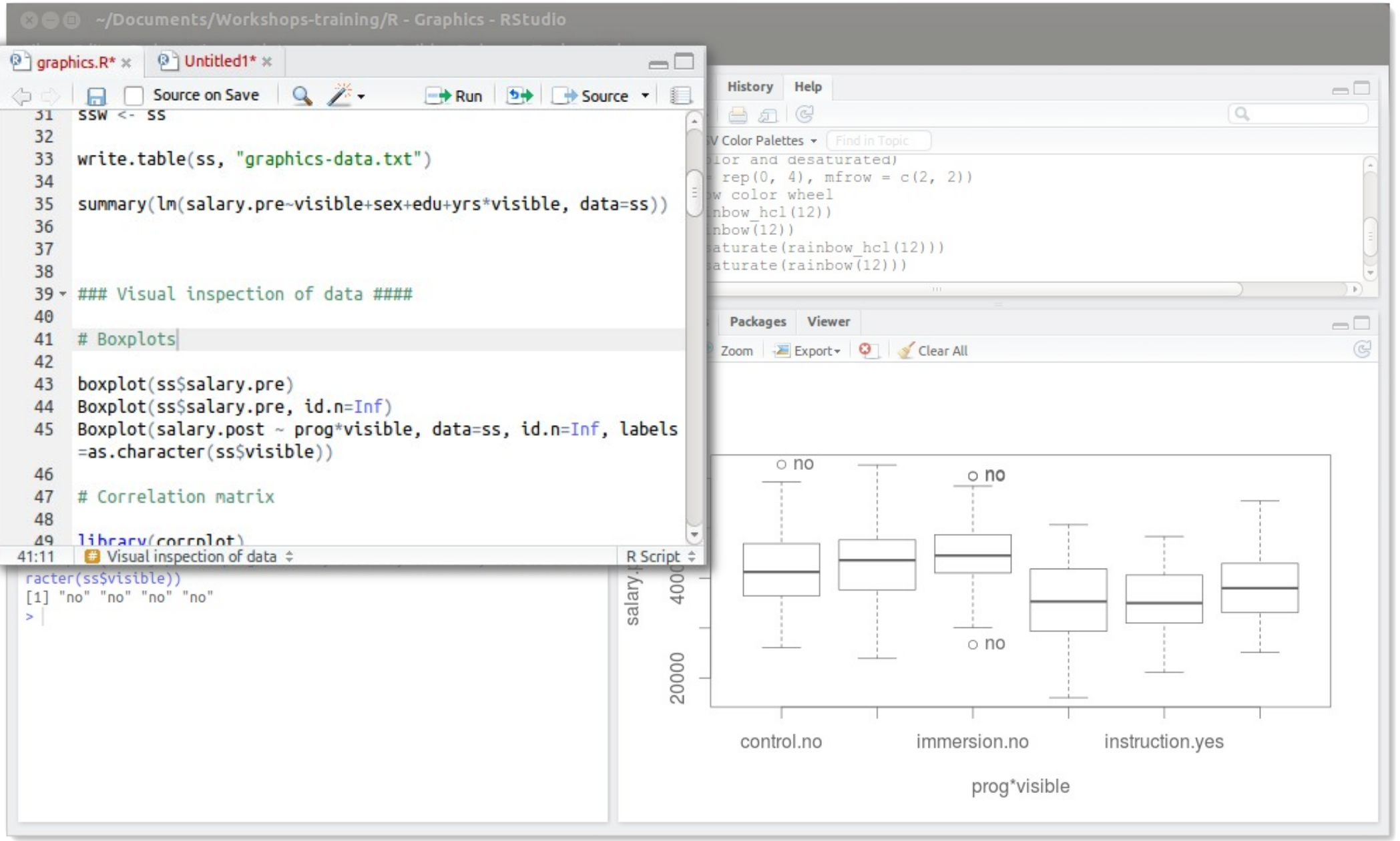


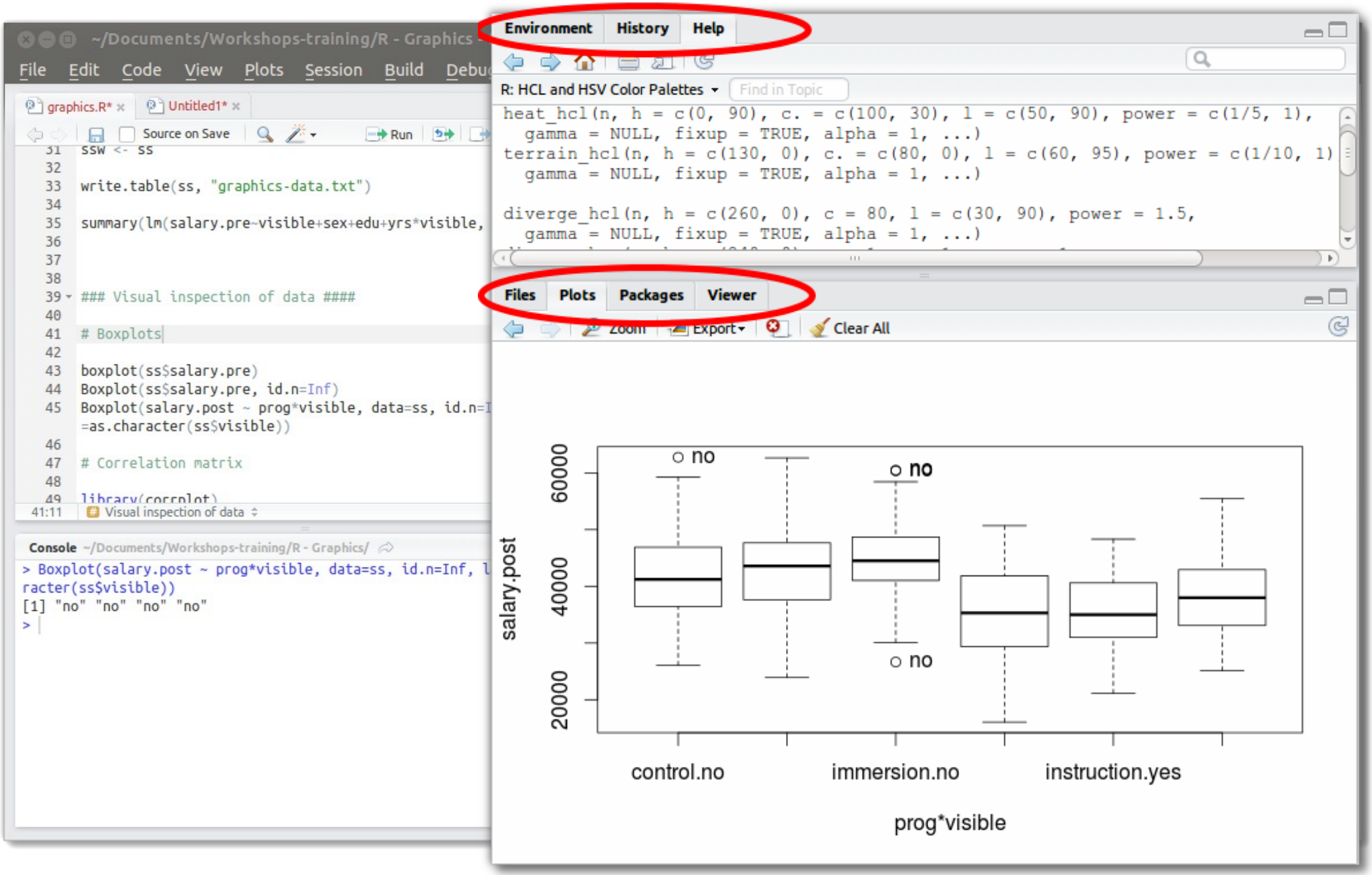
Console ~/Documents/Workshops-training/R - Graphics/

```
> Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels=as.cha
racter(ss$visible))
[1] "no" "no" "no" "no"
>
```

Console

Éditeur


~/Documents/Workshops-training/R - Graphics - RStudio
graphics.R*
Untitled1*
Source on Save
Run
Source
ssw <- ss
write.table(ss, "graphics-data.txt")
summary(lm(salary.pre~visible+sex+edu+yrs*visible, data=ss))
Visual inspection of data
# Boxplots
boxplot(ss$salary.pre)
Boxplot(ss$salary.pre, id.n=Inf)
Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels
=as.character(ss$visible))
# Correlation matrix
library(corrplot)
41:11
Visual inspection of data
R Script
racter(ss$visible))
[1] "no" "no" "no" "no"
>
History
Help
Color Palettes
Find in Topic
rep(0, 4), mfrow = c(2, 2))
color wheel
nbow_hcl(12))
nbow(12))
aturate(rainbow_hcl(12)))
aturate(rainbow(12)))
Packages
Viewer
Zoom
Export
Clear All
no
no
no
20000
control.no
immersion.no
instruction.yes
prog*visible



Vantaux supplémentaires

~/Documents/Workshops-training/R - Graphics - RStudio

File Edit Code View Plots Session Build Debug Tools Help

Import Dataset
Install Packages...
Check for Package Updates...
Version Control
Shell...
Project Options...
Global Options...

graphics.R* Untitled1*

Source on Save Run

```
ssw <- ss

write.table(ss, "graphics-data.txt")

summary(lm(salary.pre~visible+sex+edu+yrs*visible, dat

### Visual inspection of data ####

# Boxplots

boxplot(ss$salary.pre)
Boxplot(ss$salary.pre, id.n=Inf)
Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels
=as.character(ss$visible))

# Correlation matrix

library(corrplot)
```

41:11 Visual inspection of data R Script

Console ~/Documents/Workshops-training/R - Graphics/

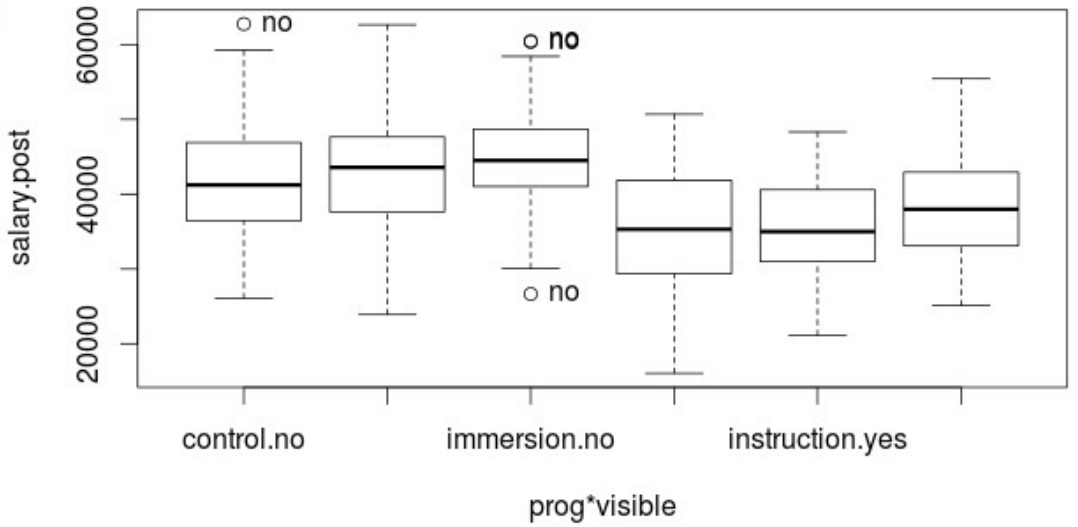
```
> Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels=as.cha
racter(ss$visible))
[1] "no" "no" "no" "no"
>
```

Help

```
esaturated)
), mfrow = c(2, 2))
heel
12))

ainbow_hcl(12)))
ainbow(12)))
```

Files Plots Packages Viewer

Zoom Export Clear All



~/Documents/Workshops-training/R - Graphics - RStudio

File Edit Code View Plots Session Build Debug Tools Help

Import Dataset
Install Packages...
Check for Package Updates...
Version Control
Shell...
Project Options...
Global Options...

graphics.R* Untitled1*

Source on Save Run

```
ssw <- ss

write.table(ss, "graphics-data.txt")

summary(lm(salary.pre~visible+sex+edu+yrs*visible, dat

### Visual inspection of data ####

# Boxplots

boxplot(ss$salary.pre)
Boxplot(ss$salary.pre, id.n=Inf)
Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels
=as.character(ss$visible))

# Correlation matrix

library(corrplot)
```

41:11 Visual inspection of data R Script

Console ~/Documents/Workshops-training/R - Graphics/

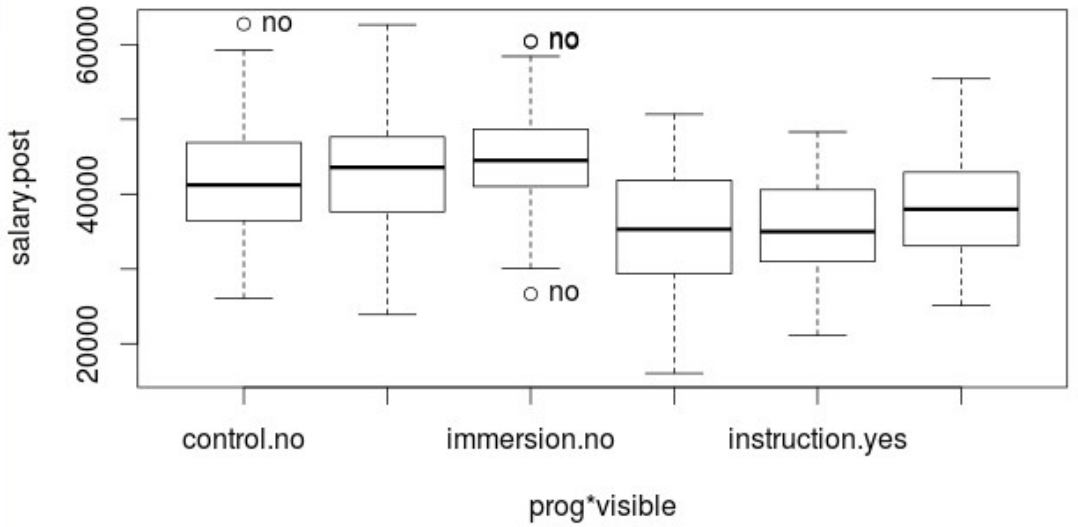
```
> Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels=as.cha
racter(ss$visible))
[1] "no" "no" "no" "no"
>
```

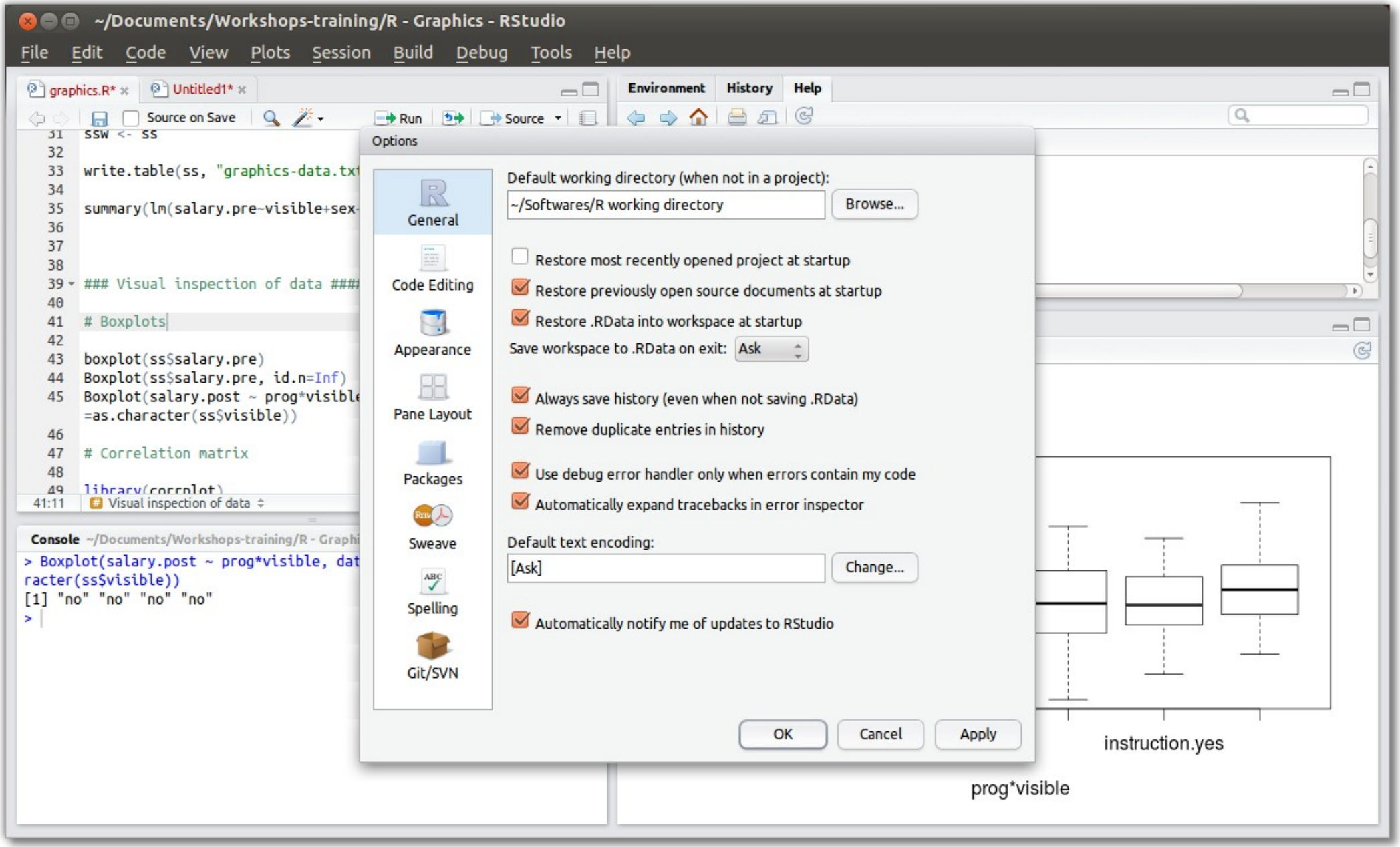
Help

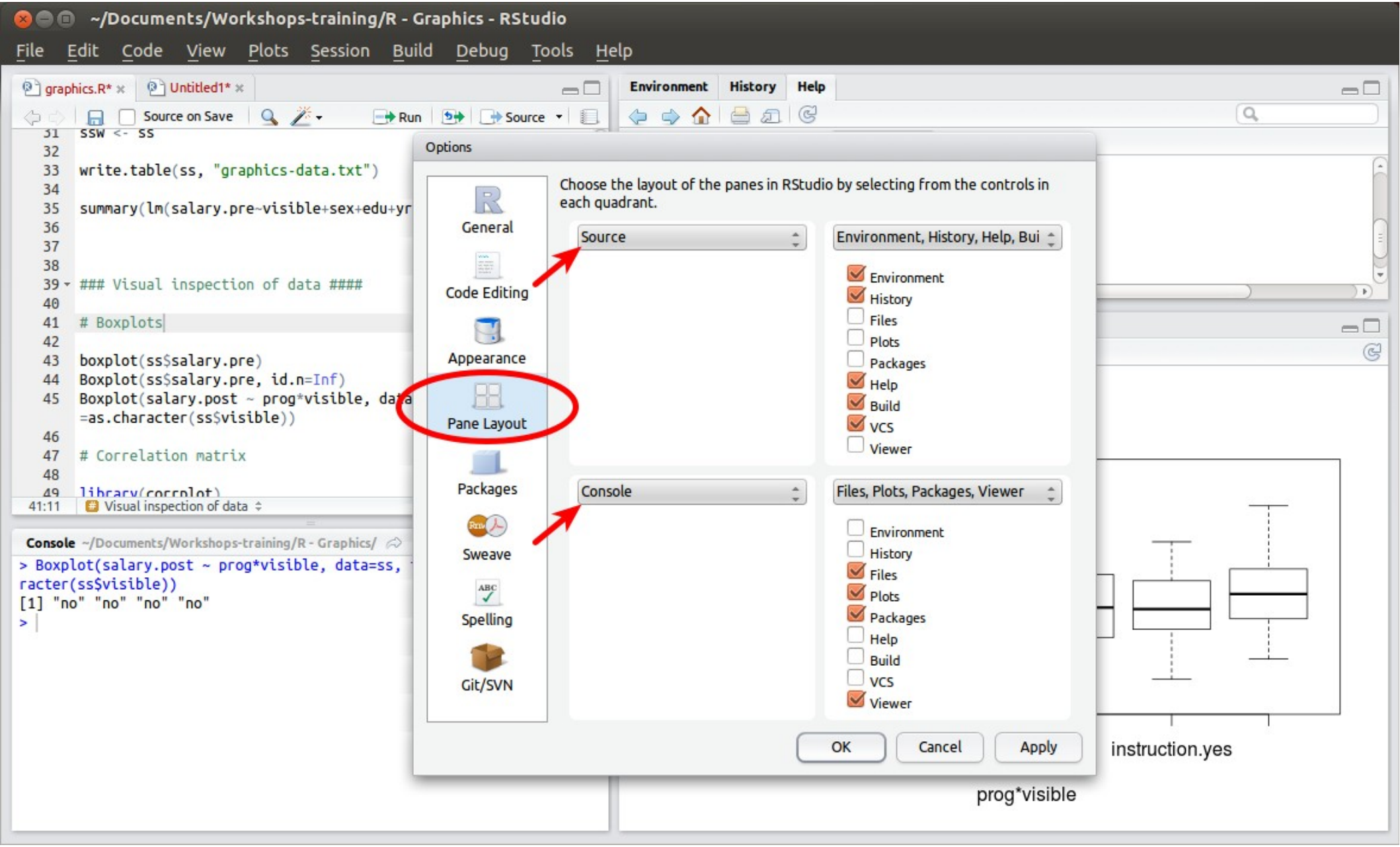
```
esaturated)
), mfrow = c(2, 2))
heel
12))

ainbow_hcl(12)))
ainbow(12)))
```

Files Plots Packages Viewer

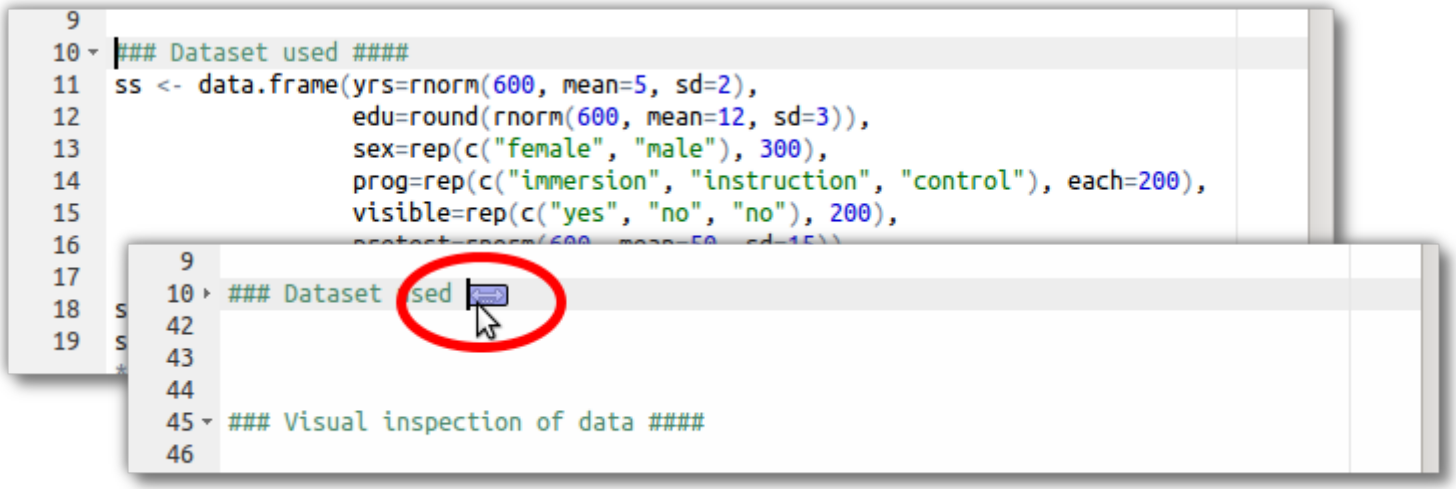
Zoom Export Clear All




~/Documents/Workshops-training/R - Graphics - RStudio
File Edit Code View Plots Session Build Debug Tools Help
graphics.R*
Untitled1*
Source on Save
Run
Source
write.table(ss, "graphics-data.txt
summary(lm(salary.pre~visible+sex
Visual inspection of data
# Boxplots
boxplot(ss$salary.pre)
Boxplot(ss$salary.pre, id.n=Inf)
Boxplot(salary.post ~ prog*visible
=as.character(ss$visible))
Correlation matrix
41:11
Visual inspection of data
Console ~/Documents/Workshops-training/R - Graphi
> Boxplot(salary.post ~ prog*visible, dat
racter(ss$visible))
[1] "no" "no" "no" "no"
Environment
History
Help
Options
General
Code Editing
Appearance
Pane Layout
Packages
Sweave
Spelling
Git/SVN
Default working directory (when not in a project):
~/Softwares/R working directory
Browse...
Restore most recently opened project at startup
Restore previously open source documents at startup
Restore .RData into workspace at startup
Save workspace to .RData on exit:
Ask
Always save history (even when not saving .RData)
Remove duplicate entries in history
Use debug error handler only when errors contain my code
Automatically expand tracebacks in error inspector
Default text encoding:
[Ask]
Change...
Automatically notify me of updates to RStudio
OK
Cancel
Apply
instruction.yes
prog*visible


~/Documents/Workshops-training/R - Graphics - RStudio
File Edit Code View Plots Session Build Debug Tools Help
graphics.R*
Untitled1*
Source on Save
Run
Source
ssw <- ss
write.table(ss, "graphics-data.txt")
summary(lm(salary.pre~visible+sex+edu+yr
Visual inspection of data
# Boxplots
boxplot(ss$salary.pre)
Boxplot(ss$salary.pre, id.n=Inf)
Boxplot(salary.post ~ prog*visible, da
=as.character(ss$visible))
# Correlation matrix
library(corrplot)
41:11
Visual inspection of data
Console ~/Documents/Workshops-training/R - Graphics/
> Boxplot(salary.post ~ prog*visible, data=ss,
racter(ss$visible))
[1] "no" "no" "no" "no"
>
Environment
History
Help
Options
General
Code Editing
Appearance
Pane Layout
Packages
Sweave
Spelling
Git/SVN
Choose the layout of the panes in RStudio by selecting from the controls in each quadrant.
Source
Environment, History, Help, Bui
Environment
History
Files
Plots
Packages
Help
Build
VCS
Viewer
Console
Files, Plots, Packages, Viewer
Environment
History
Files
Plots
Packages
Help
Build
VCS
Viewer
OK
Cancel
Apply
instruction.yes
prog*visible

Réduire le code




```
# Possibilité 1 ------------------------
# Possibilité 2 ========================
# Possibilité 3 ########################
```

Raccourcis

Dans l'éditeur:

Ctrl + Entrée: exécute code dans la console

Ctrl + 2: déplace le curseur dans la console

Dans la console:

Ctrl + 1: déplace le curseur dans l'éditeur

Flèche vers le haut: recouvre la commande précédente

~/Documents/Workshops-training/R - Graphics - RStudio

File Edit Code View Plots Session Build Debug Tools Help

graphics.R* Untitled1*

Source on Save Run Source

```
31  ssw <- ss
32
33  write.table(ss, "graphics-data.txt")
34
35  summary(lm(salary.pre~visible+sex+edu+yrs*visible, data=ss))
36
37
38
39  ### Visual inspection of data ####
40
41  # Boxplots
42
43  boxplot(ss$salary.pre)
44  Boxplot(ss$salary.pre, id.n=Inf)
```

Environment History Help

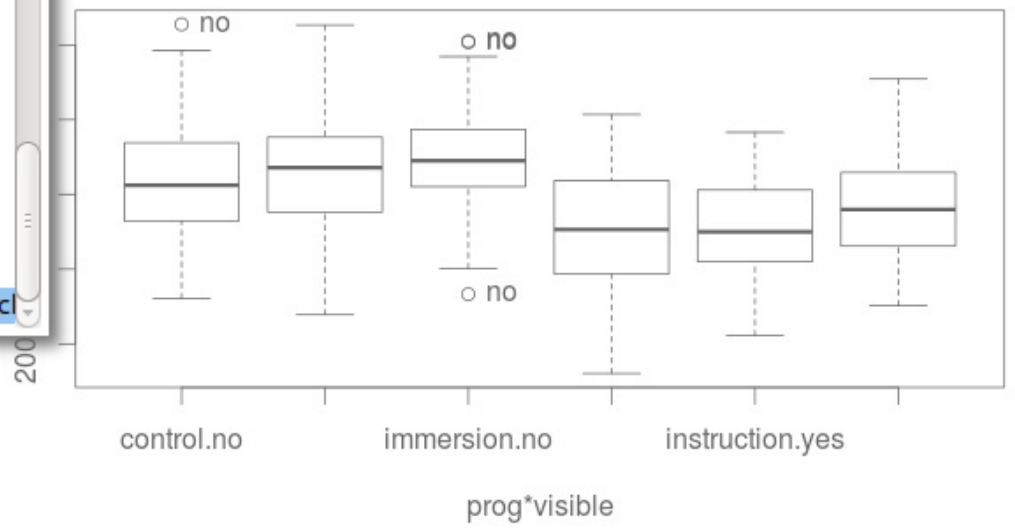
R: HCL and HSV Color Palettes

```
## (in color and desaturated)
par(mar = rep(0, 4), mfrow = c(2, 2))
## rainbow color wheel
wheel(rainbow_hcl(12))
wheel(rainbow(12))
wheel(desaturate(rainbow_hcl(12)))
wheel(desaturate(rainbow(12)))
```

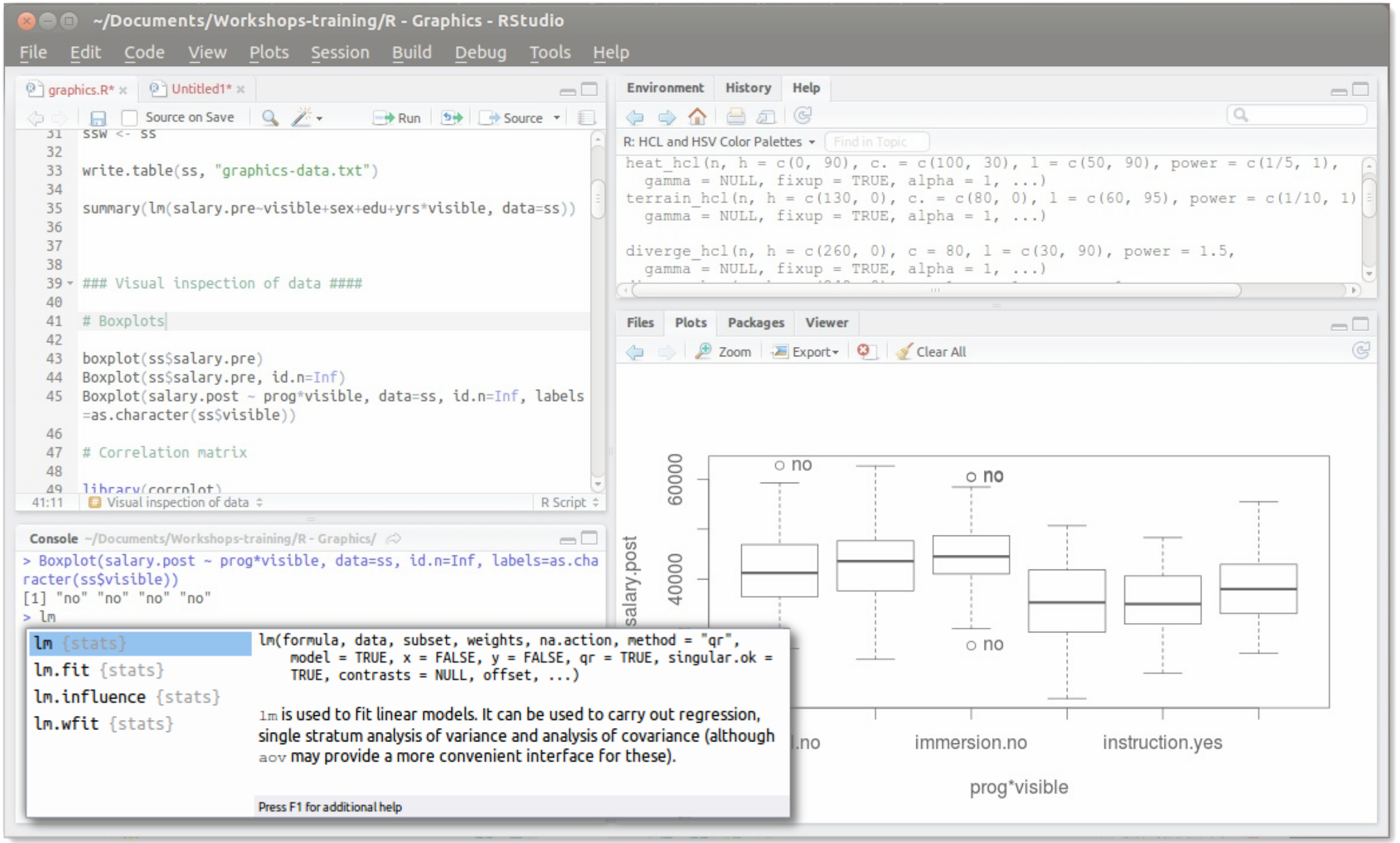
Files Plots Packages Viewer

Zoom Export Clear All

```
stat_bin(geom="line")
ggplot(ss, aes(x=posttest))+
stat_bin(geom="bar")
ggplot(diamonds, aes(x=price))
geom_histogram(aes(y = ..density..))
geom_histogram(aes(colour = ..count..))
ggplot(diamonds, aes(x=price)) +
geom_histogram(aes(fill = ..count..))
library(car)
Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels=as.cl
```



[Ctrl + ↑]


~/Documents/Workshops-training/R - Graphics - RStudio
File Edit Code View Plots Session Build Debug Tools Help
graphics.R*
Untitled1*
Source on Save
Run
Source
ssw <- ss
write.table(ss, "graphics-data.txt")
summary(lm(salary.pre~visible+sex+edu+yrs*visible, data=ss))
Visual inspection of data
# Boxplots
boxplot(ss$salary.pre)
Boxplot(ss$salary.pre, id.n=Inf)
Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels=as.character(ss$visible))
# Correlation matrix
library(corrplot)
41:11
Visual inspection of data
R Script
Console ~/Documents/Workshops-training/R - Graphics/
> Boxplot(salary.post ~ prog*visible, data=ss, id.n=Inf, labels=as.character(ss$visible))
[1] "no" "no" "no" "no"
> lm
lm {stats}
lm.fit {stats}
lm.influence {stats}
lm.wfit {stats}
lm(formula, data, subset, weights, na.action, method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE, singular.ok = TRUE, contrasts = NULL, offset, ...)
lm is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of covariance (although aov may provide a more convenient interface for these).
Press F1 for additional help
Environment History Help
R: HCL and HSV Color Palettes
Find in Topic
heat_hcl(n, h = c(0, 90), c. = c(100, 30), l = c(50, 90), power = c(1/5, 1), gamma = NULL, fixup = TRUE, alpha = 1, ...)
terrain_hcl(n, h = c(130, 0), c. = c(80, 0), l = c(60, 95), power = c(1/10, 1) gamma = NULL, fixup = TRUE, alpha = 1, ...)
diverge_hcl(n, h = c(260, 0), c = 80, l = c(30, 90), power = 1.5, gamma = NULL, fixup = TRUE, alpha = 1, ...)
Files Plots Packages Viewer
Zoom
Export
Clear All
salary.post
60000
40000
no
no
no
immersion.no
instruction.yes
prog*visible

[Tab]

Jeu de données

```
> head(ss)
```

yrs	edu	sex	prog	visible	pretest	salary.pre	salary.post	posttest
4.6	11	female	immersion	yes	34	29556	35657	60
2.8	10	male	immersion	no	84	38183	44390	100
9.5	10	female	immersion	no	33	41334	48468	59
4.3	14	male	immersion	yes	34	34837	40429	60
7.6	6	female	immersion	no	42	30713	36917	67
6.1	14	male	immersion	no	46	43973	49498	72

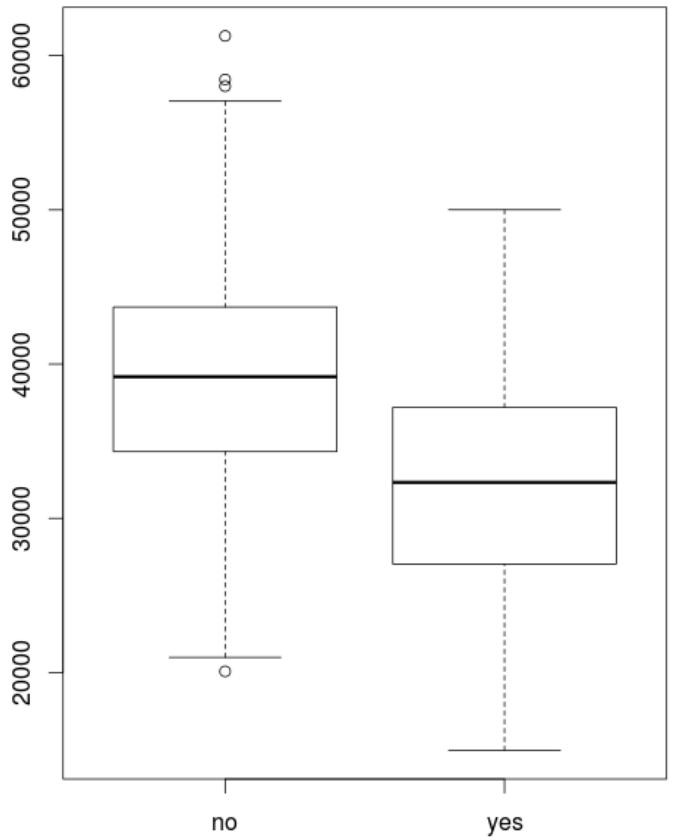
Exploration & diagnostiques

car

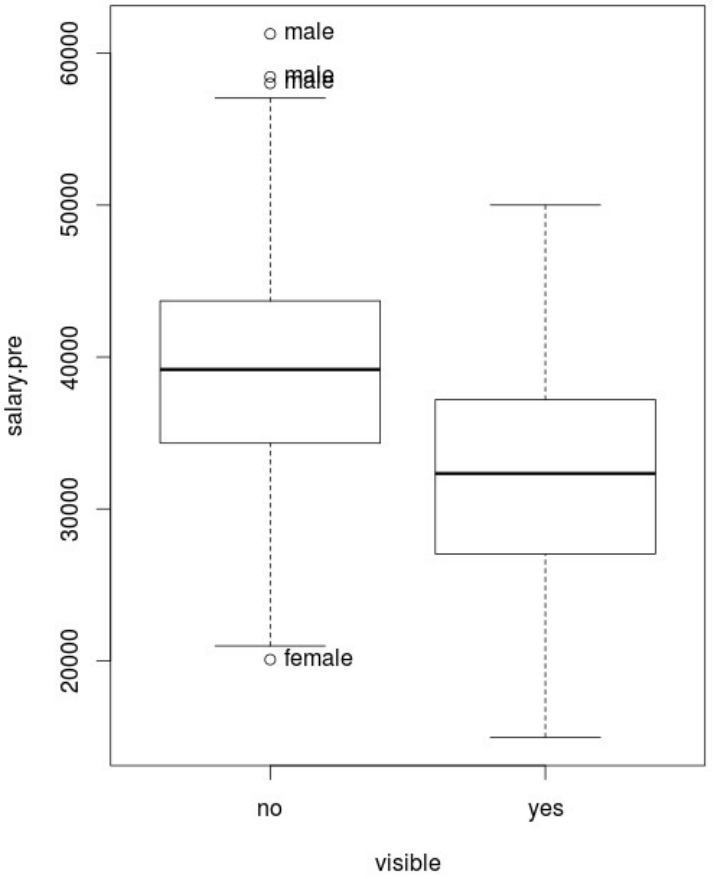
(John Fox)

1) Distributions

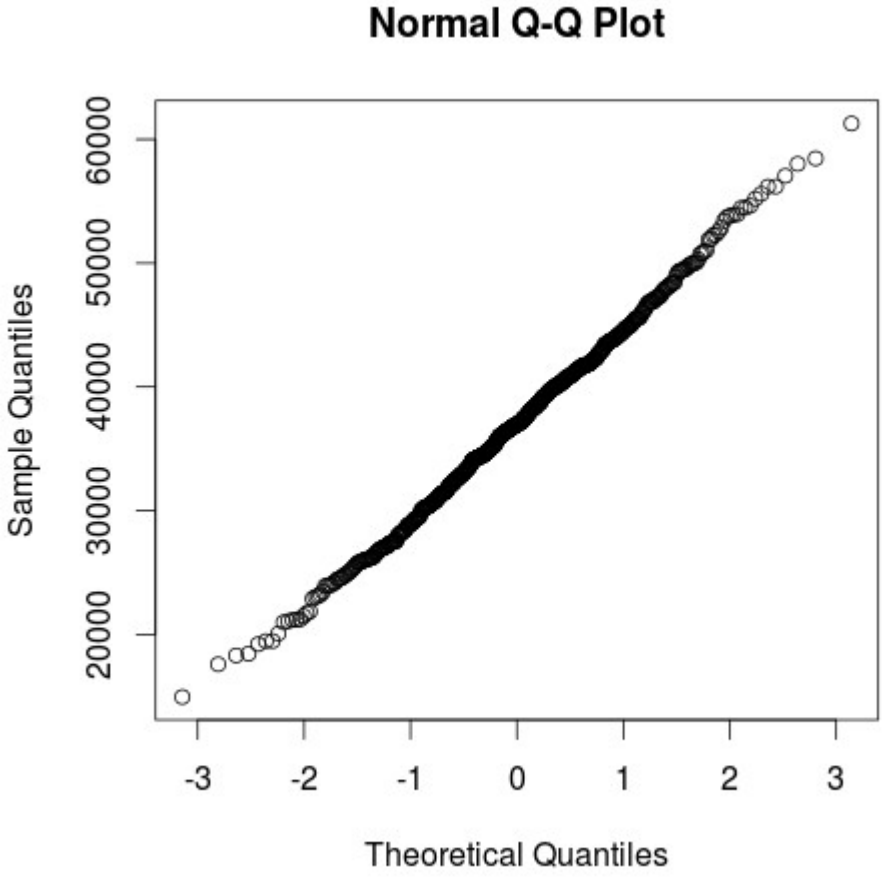
```
> boxplot(salary.pre ~ visible,
          data=ss)
```



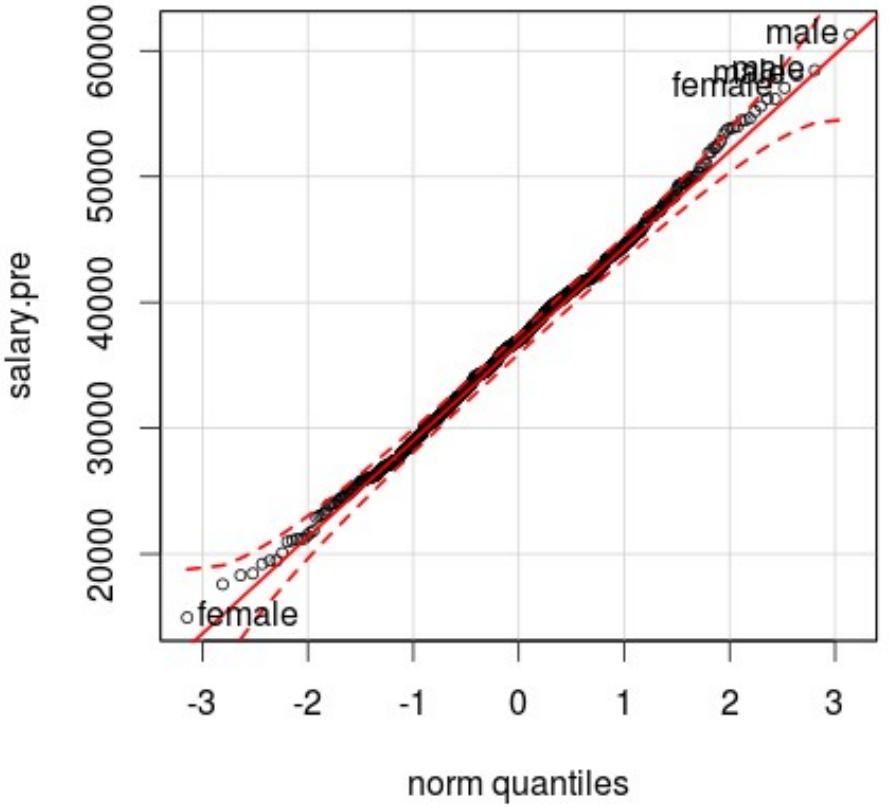

```
> Boxplot(salary.pre ~ visible,
          data=ss,
          id.n=Inf,
          labels=as.character(ss$sex))
```



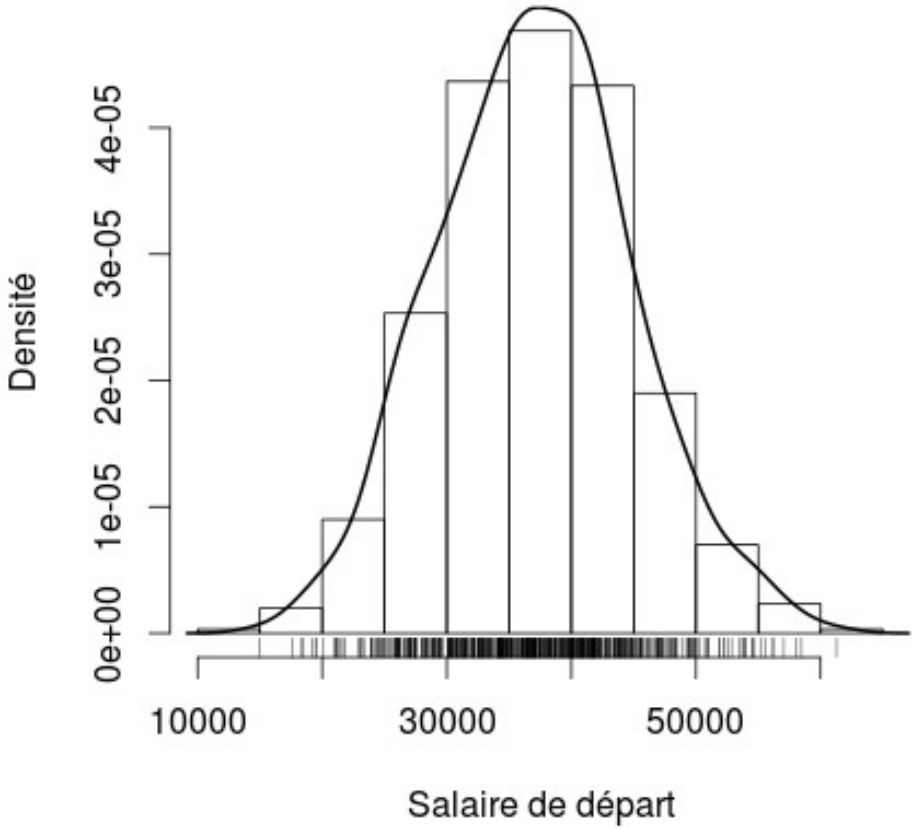
```
> qqnorm(ss$salary.pre)
```




```
> with(ss, qqPlot(salary.pre,
      labels=sex, id.n=5, cex=.7))
```

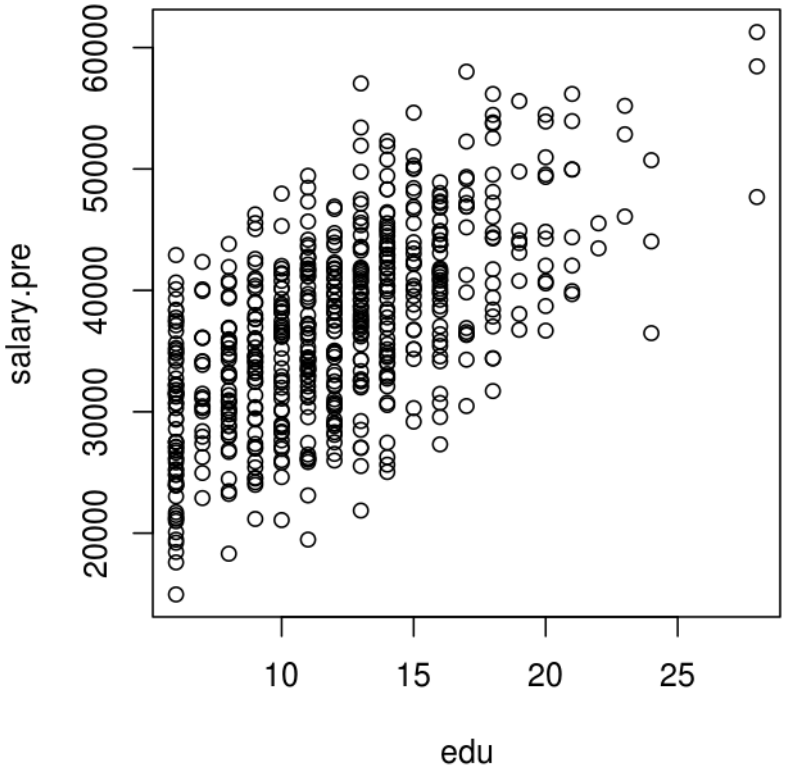


```
> with(ss, {
    hist(salary.pre, freq=F, ylab="Densité", main=NULL,
           xlab="Salaire de départ")
    lines(density(salary.pre), lwd=2)
    rug(salary.pre)
  })
```

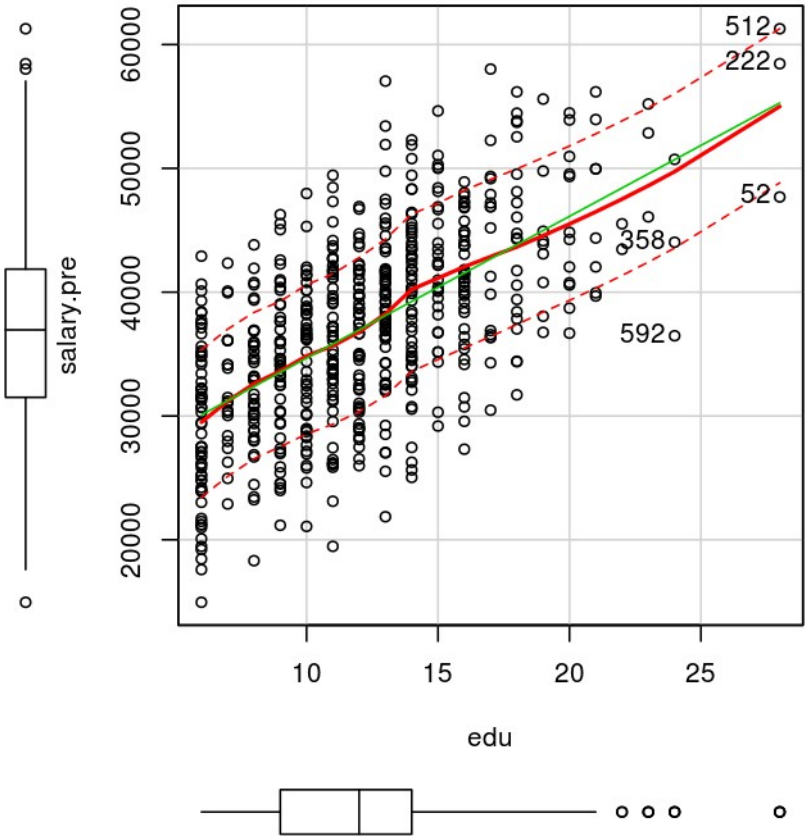


2) Relations

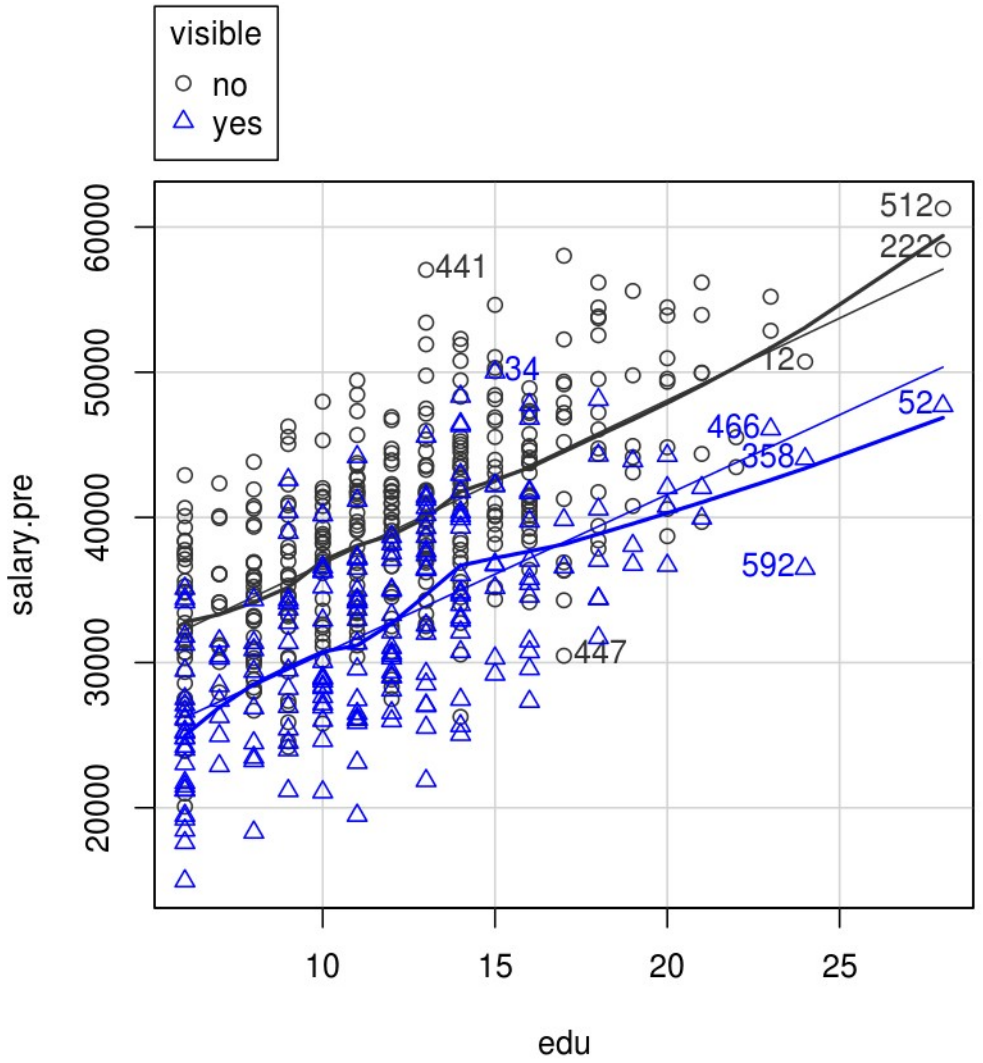
```
> plot(salary.pre~edu,data=ss)
```



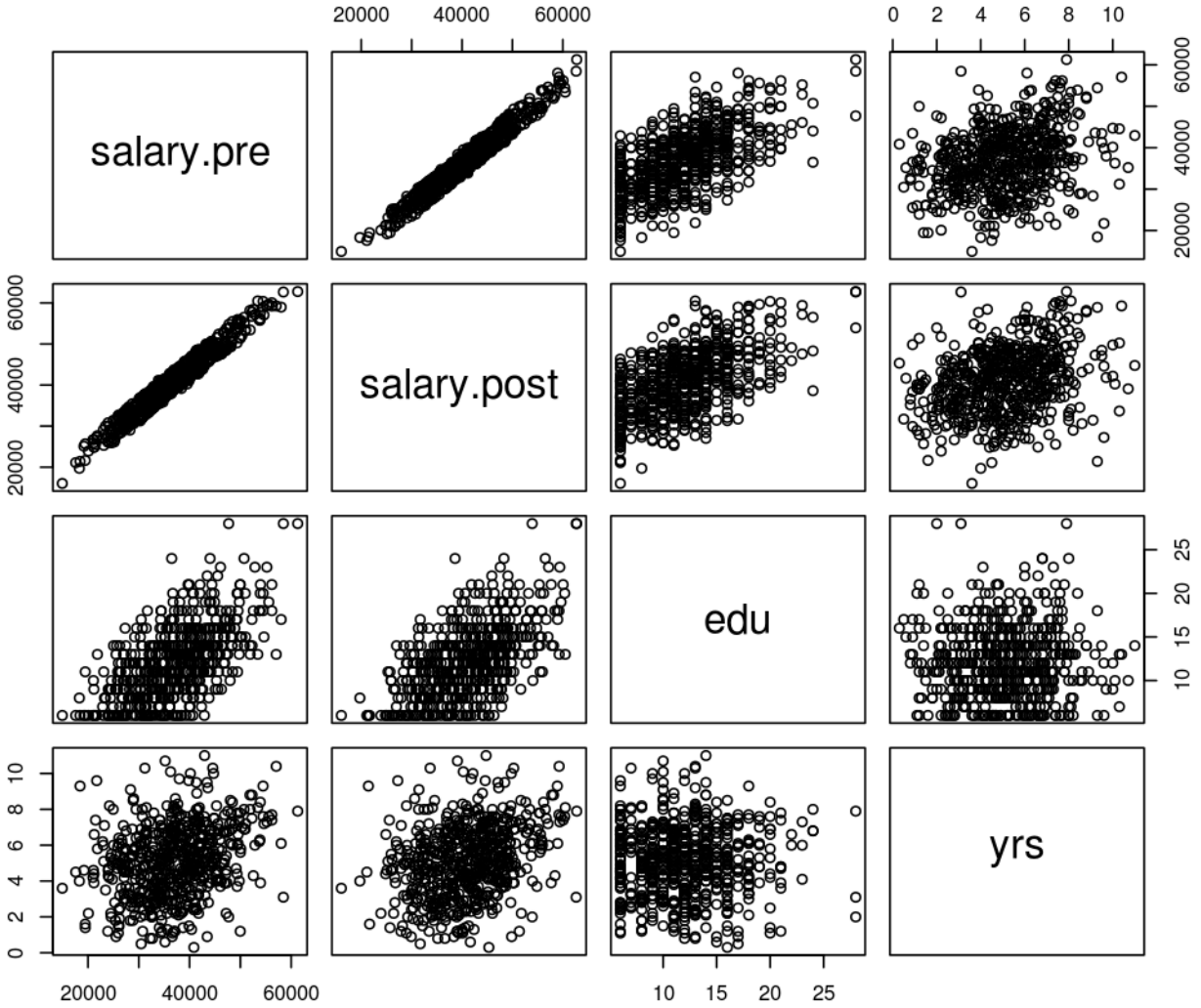

```
> scatterplot(salary.pre~edu,
                data=ss, id.n=5 )
```



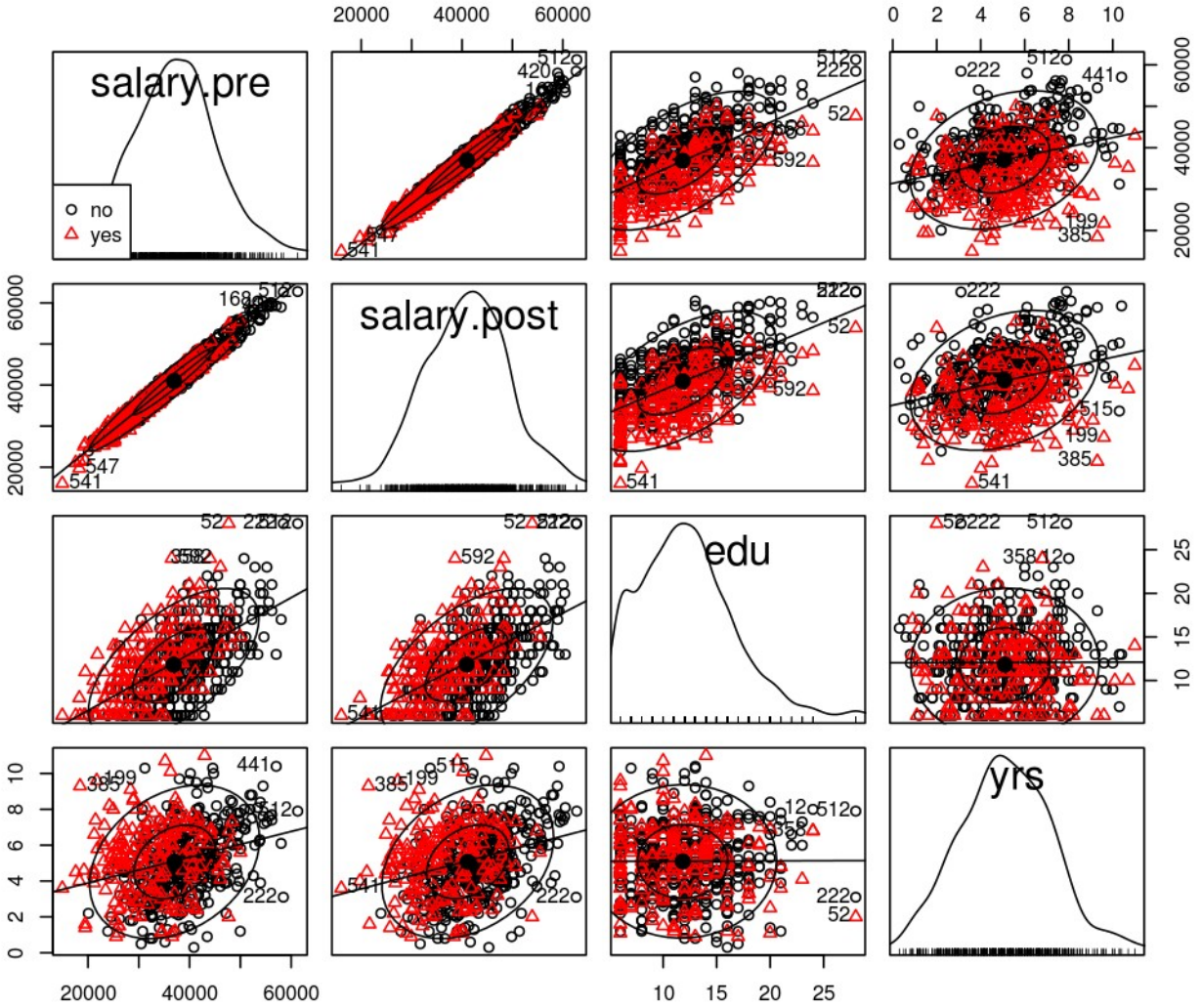
```
> scatterplot(salary.pre~edu | visible, data=ss, id.n=5 )
```



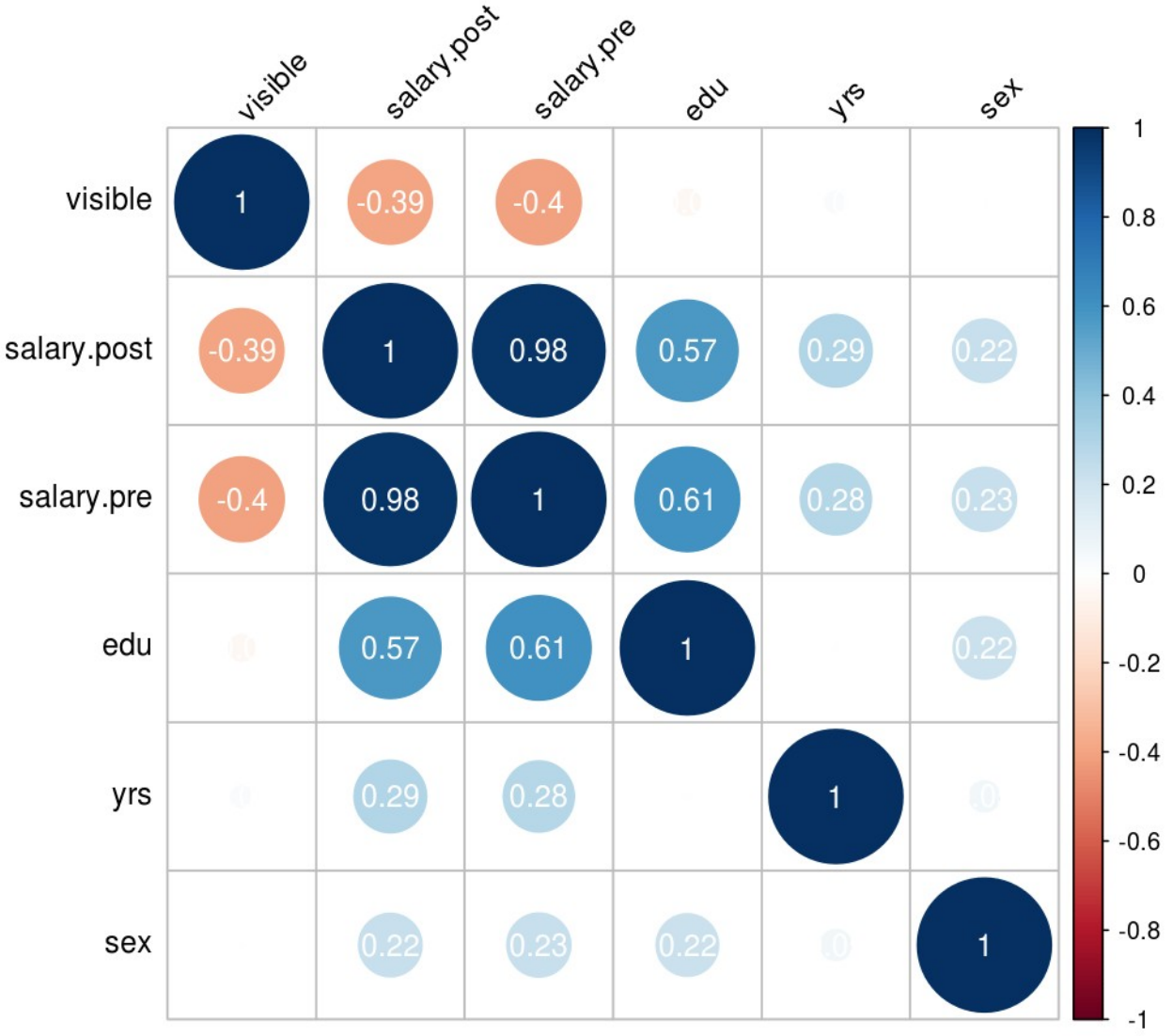
```
> pairs(ss[,c("salary.pre", "salary.post", "edu", "yrs")])
```



```
> scatterplotMatrix(~salary.pre+ salary.post+ edu+ yrs, data=ss,
    smooth=F, diagonal="density", ellipse=T, id.n=5,
    groups=ss$visible)
```



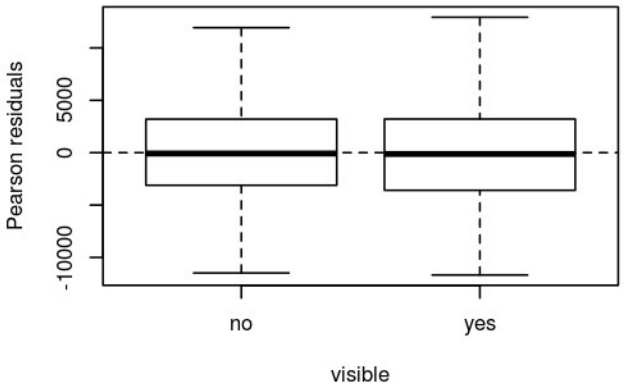
```
> corrplot(cor(mapply(as.numeric, ss[,c("yrs", "salary.pre", "edu",
          "salary.post", "visible", "sex")])),
          addCoef.col="grey80", order="AOE",
          tl.col="black", tl.srt=45)
```

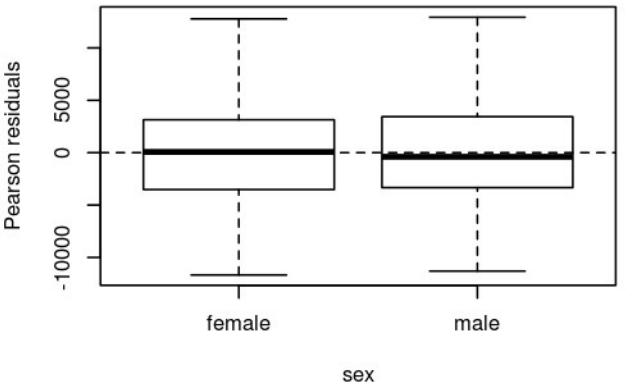


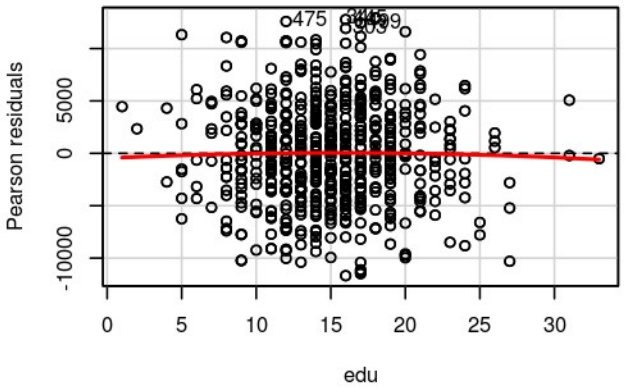
3) Diagnostiques

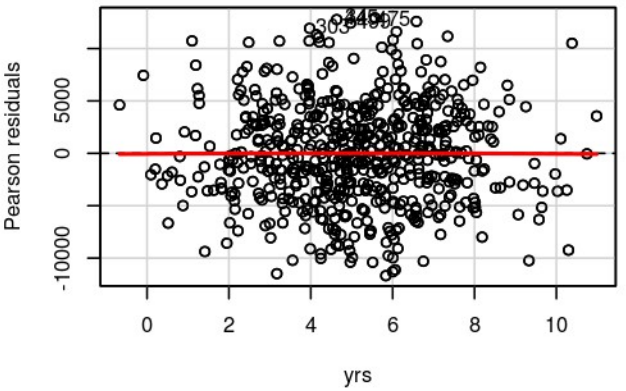
```
mod <- lm(salary.pre~sex+edu+yrs*visible,data=ss)
```

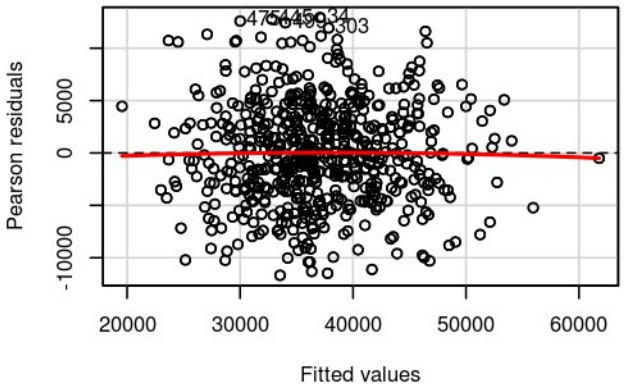
```
> residualPlots(mod)
```



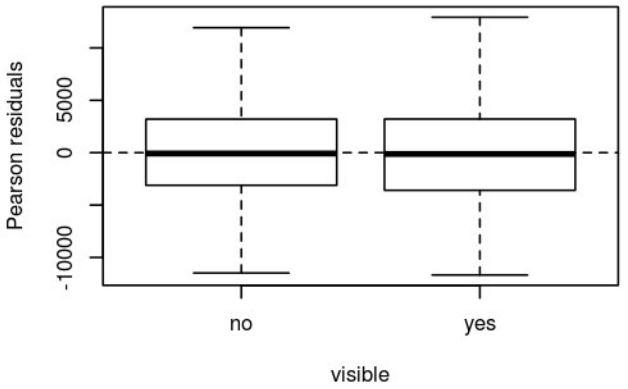


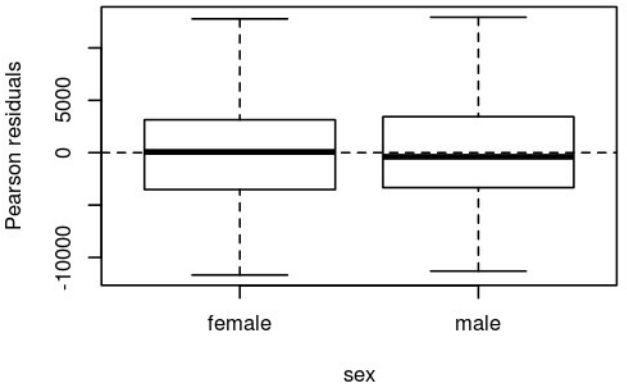


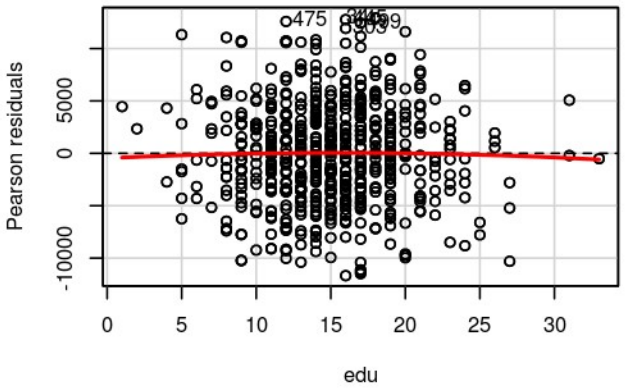


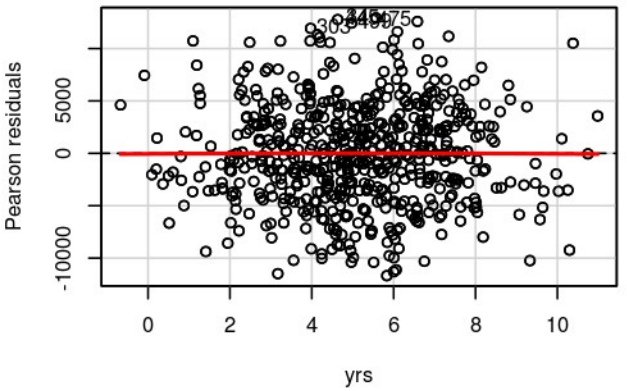


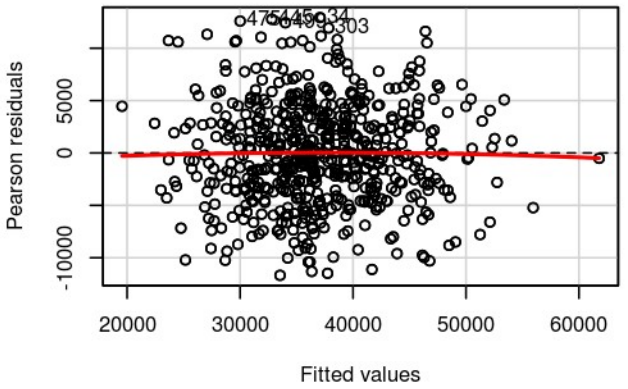
```
> residualPlots(mod)
```





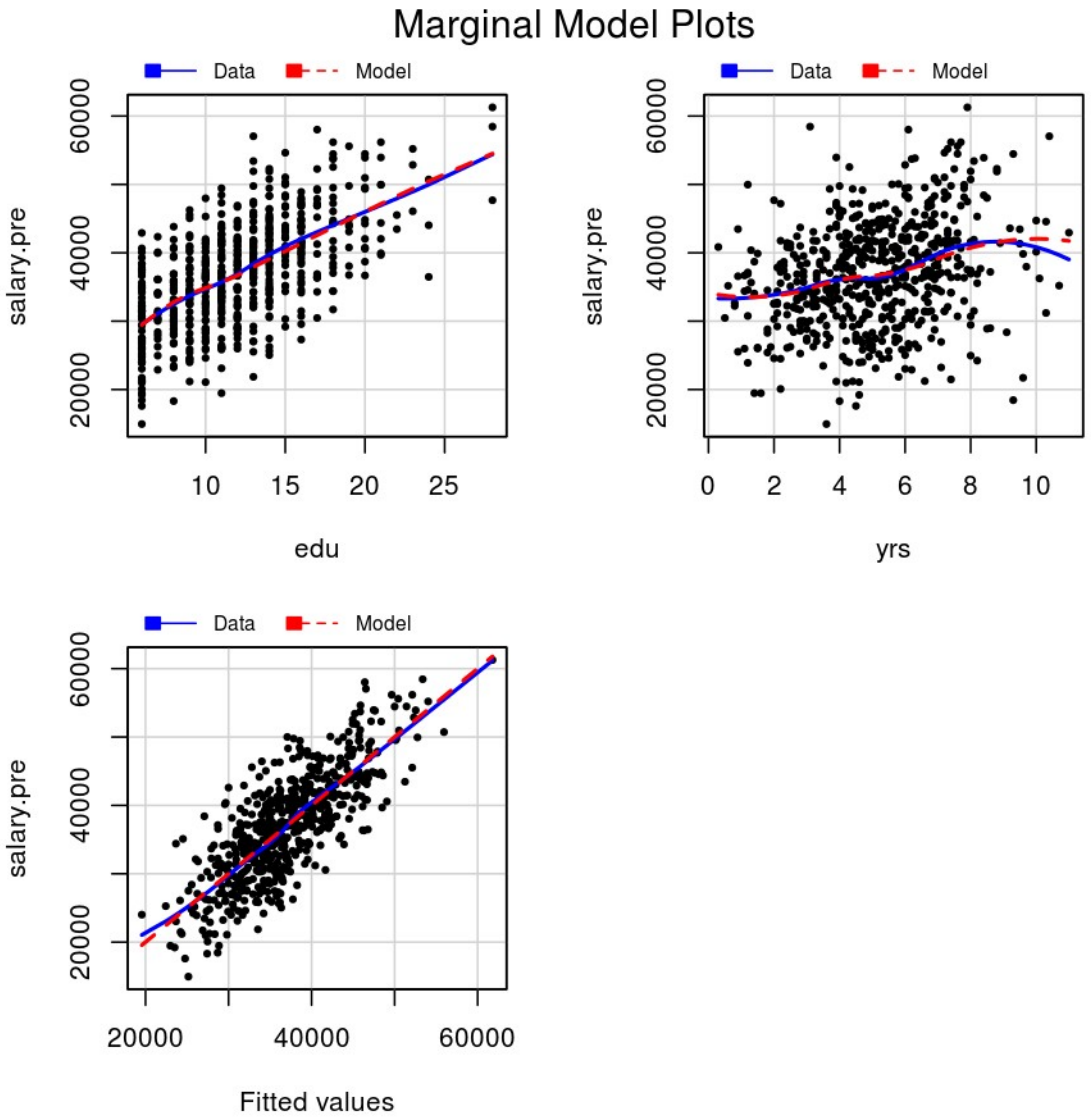




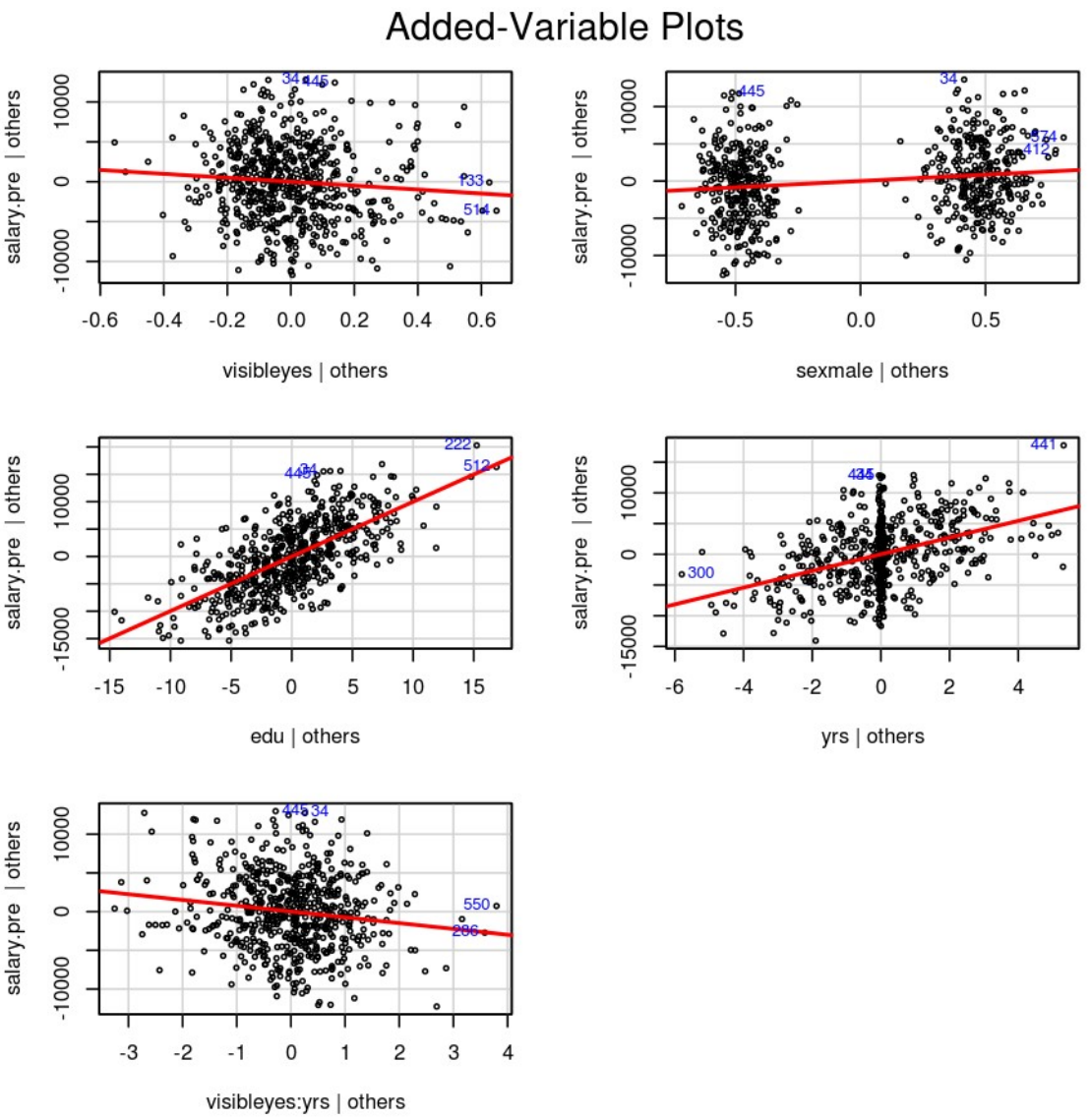


	Test stat	Pr(>|t|)
visible	NA	NA
sex	NA	NA
edu	-.342	.732
yrs	-.064	.949
Tukey	-.269	.799

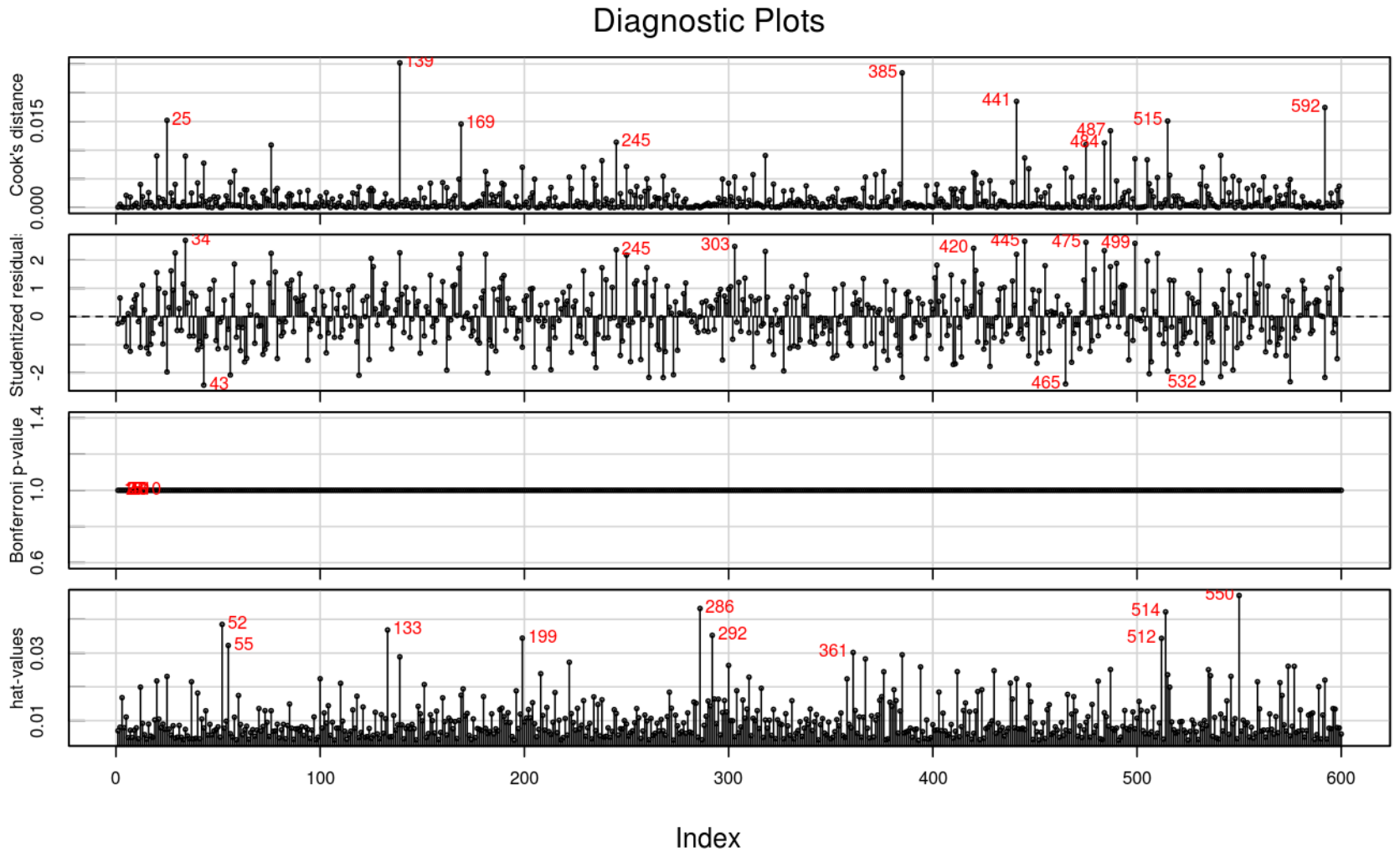
```
> marginalModelPlots(mod, cex=.5, pch=19)
```



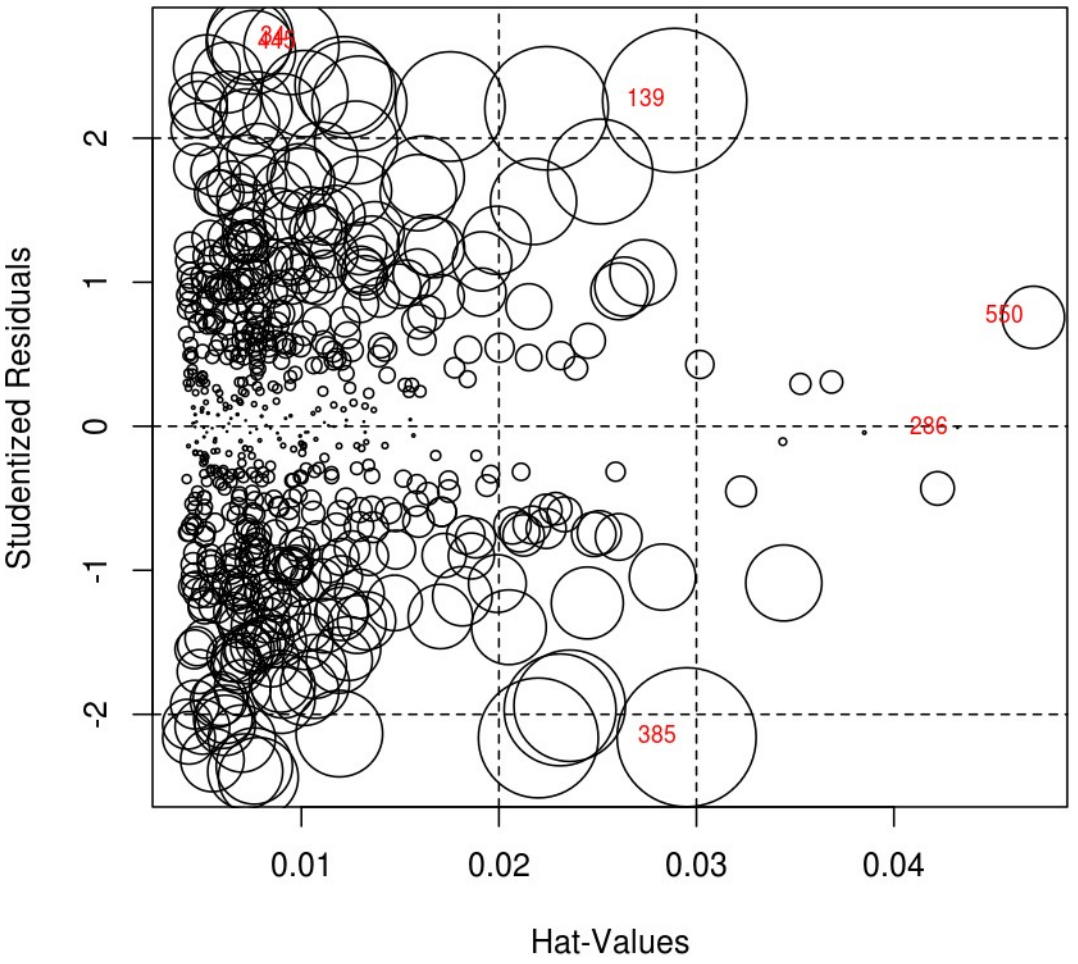
```
> avPlots(mod, id.n=2, id.cex=.8, id.col="blue", cex=.5)
```



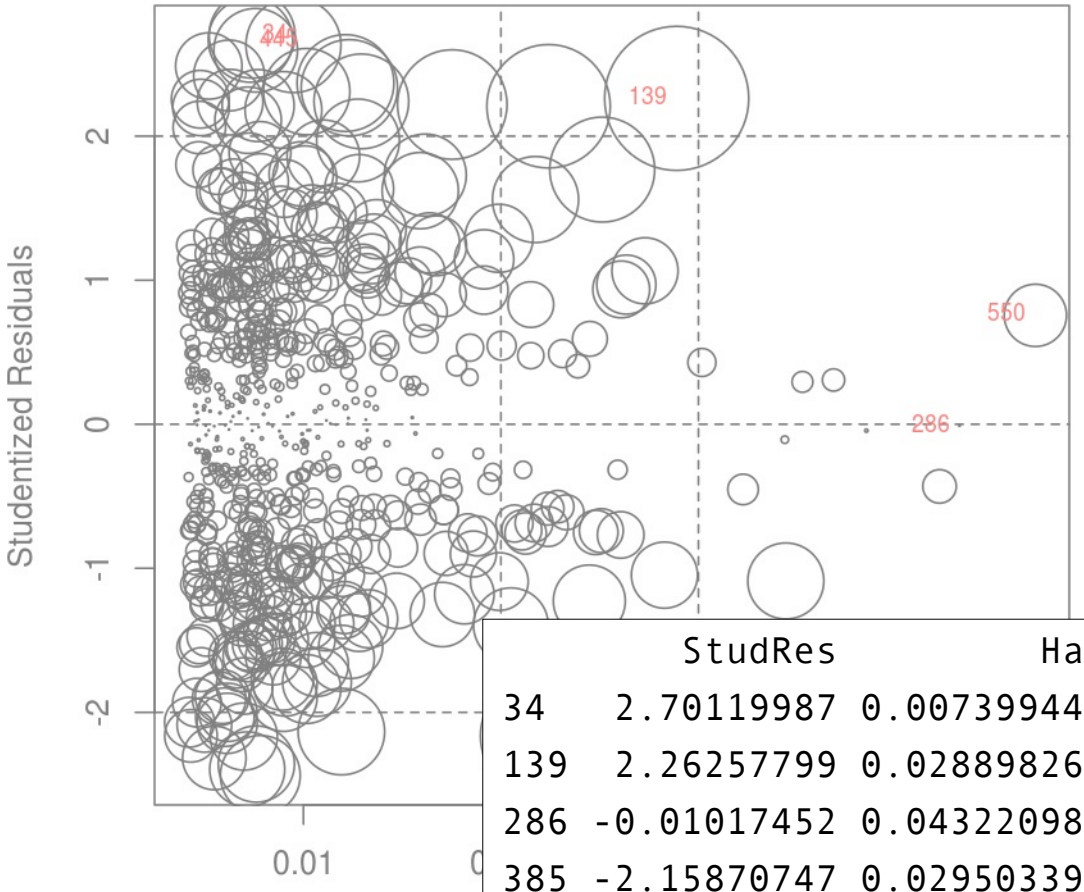
```
> influenceIndexPlot(mod, cex=.5, id.n=10, id.col="red")
```



```
> influencePlot(mod, id.n=2, id.col="red", id.cex=.7)
```



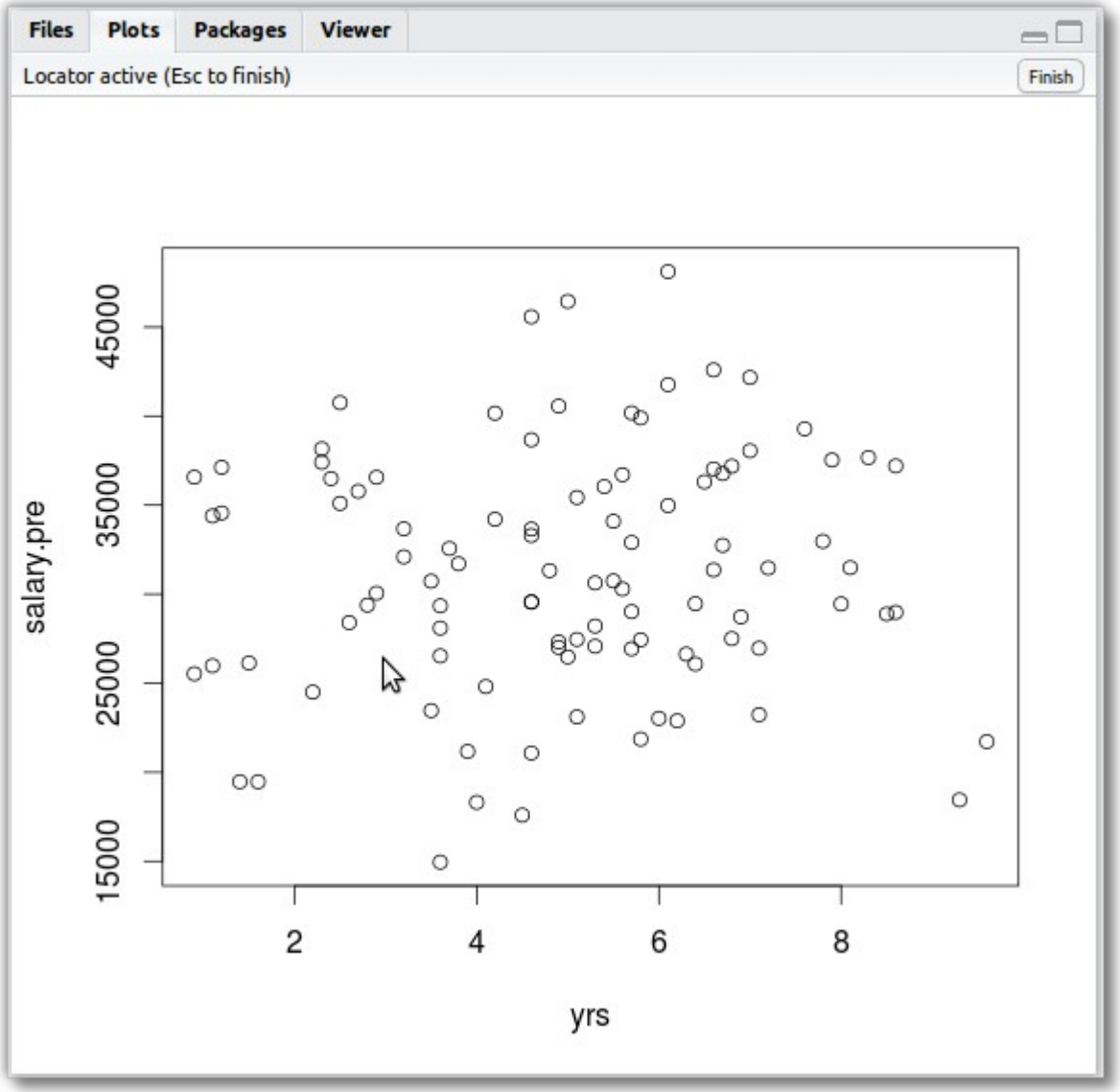
```
> influencePlot(mod, id.n=2, id.col="red", id.cex=.7)
```



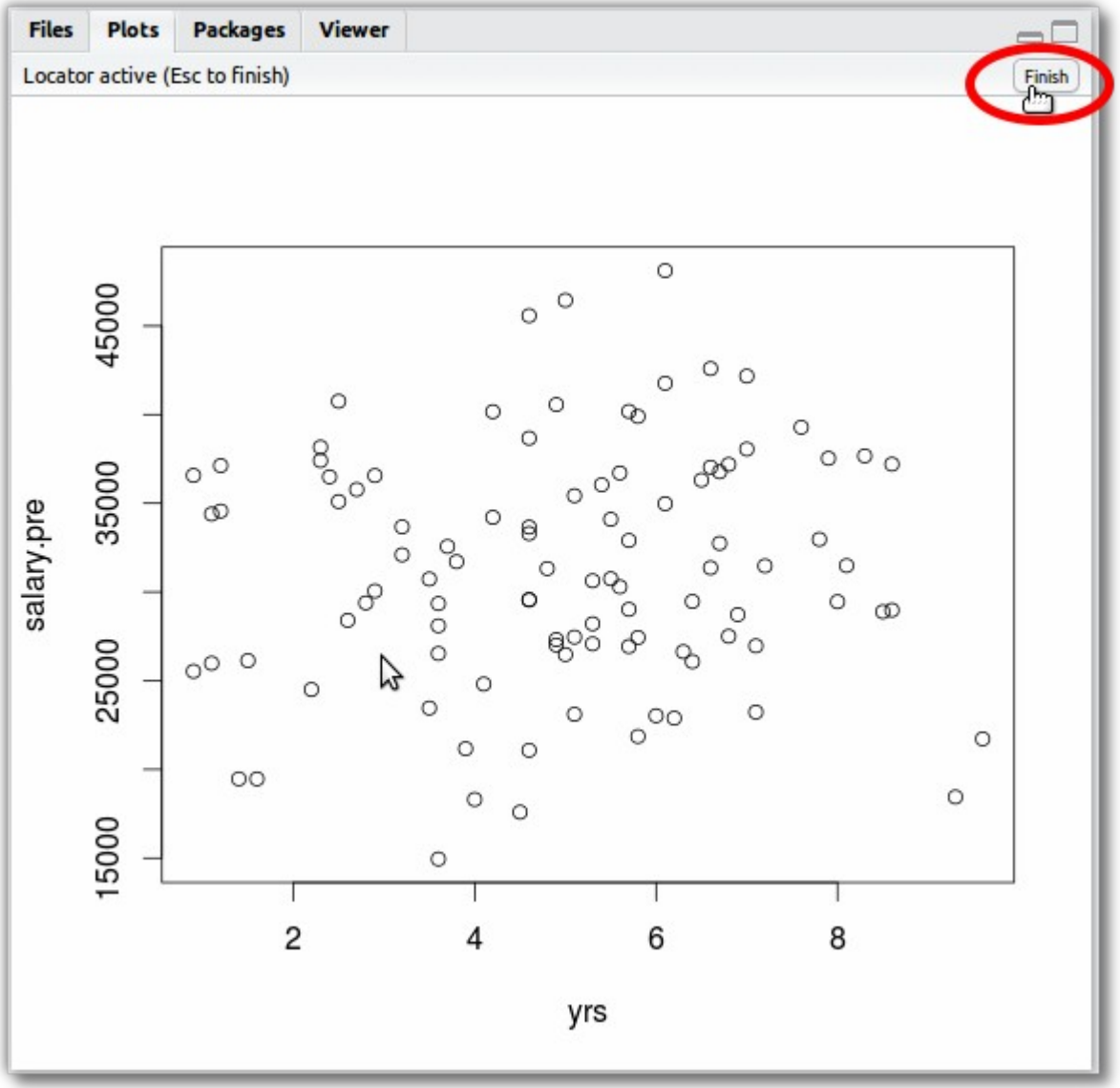
	StudRes	Hat	CookD
34	2.70119987	0.007399443	0.0947117091
139	2.26257799	0.028898260	0.1587927762
286	-0.01017452	0.043220989	0.0008835796
385	-2.15870747	0.029503399	0.1531874925
445	2.66740451	0.007319942	0.0930332907
550	0.75618242	0.047061550	0.0686290876

4) Identification

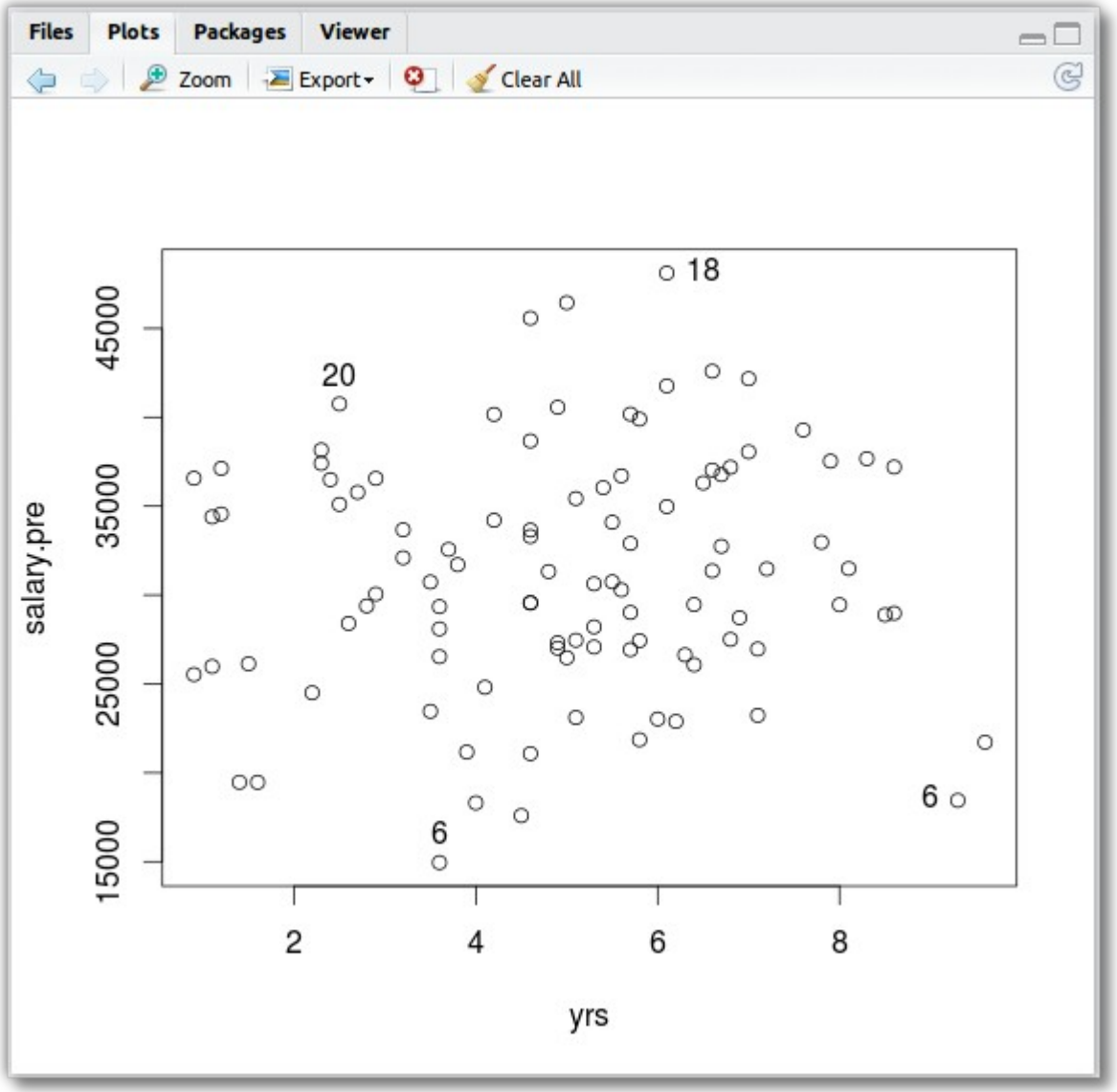
```
> with(subset(ss, sex=="female" & visible=="yes"), {
    plot(salary.pre~yrs)
    identify(yrs, salary.pre, edu)
  })
```



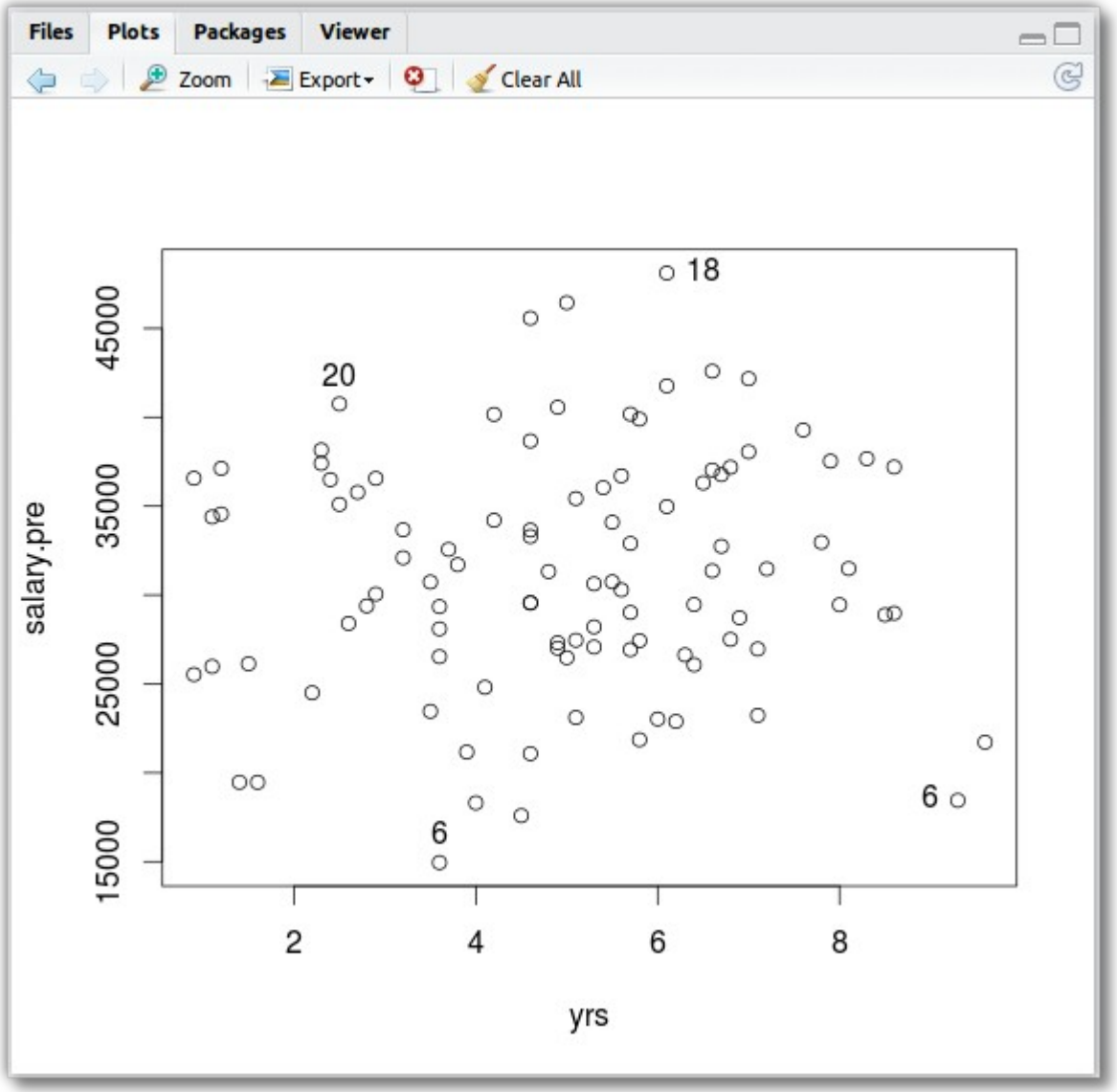
```
> with(subset(ss, sex=="female" & visible=="yes"), {
    plot(salary.pre~yrs)
    identify(yrs, salary.pre, edu)
  })
```



```
> with(subset(ss, sex=="female" & visible=="yes"), {
    plot(salary.pre~yrs)
    identify(yrs, salary.pre, edu)
  })
```



```
> with(subset(ss, sex=="female" & visible=="yes"), {
    plot(salary.pre~yrs)
    identify(yrs, salary.pre, edu)
  })
```



```
> scatterplot(salary.pre~yrs,
              data=ss,
              id.method="identify", labels=ss$edu)
```

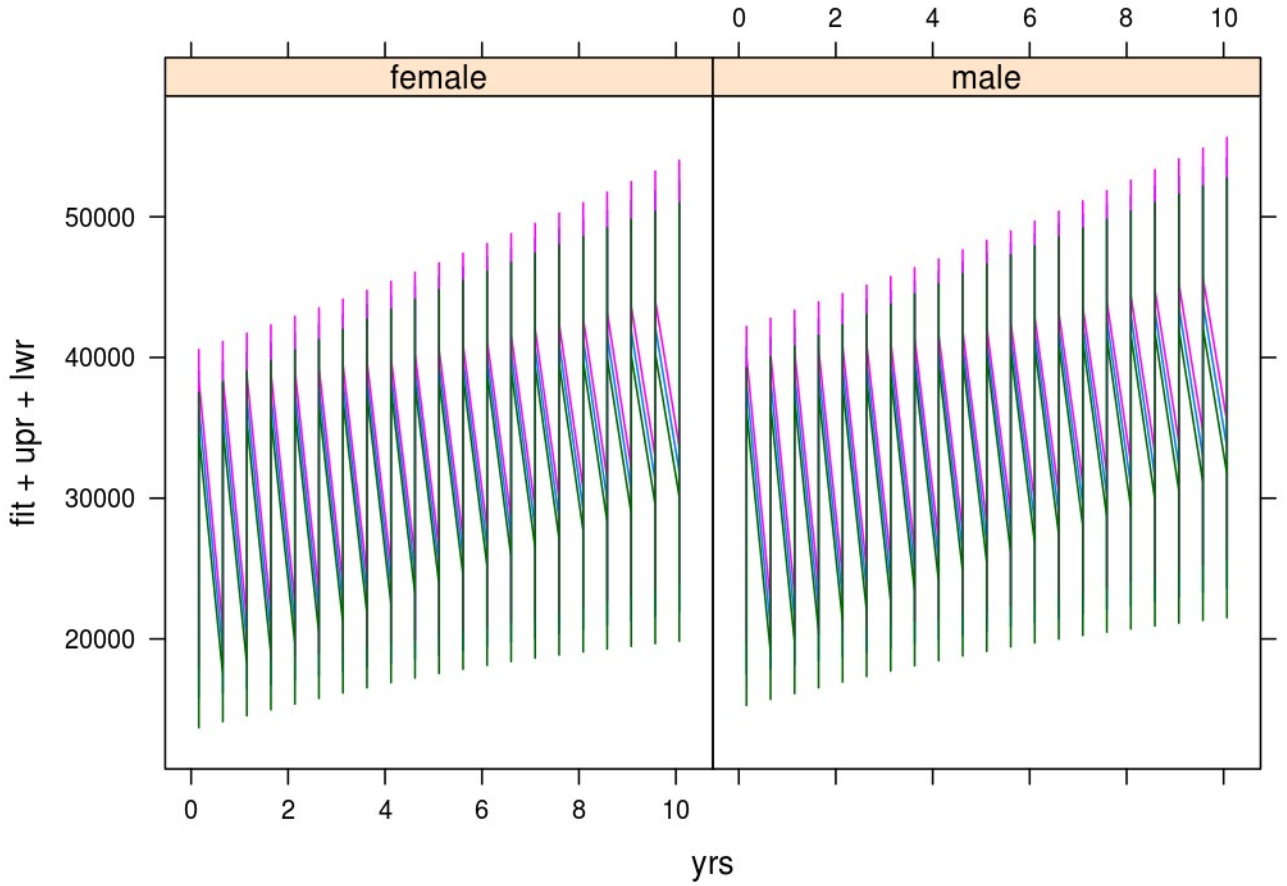
lattice

(Deepayan Sarkar)

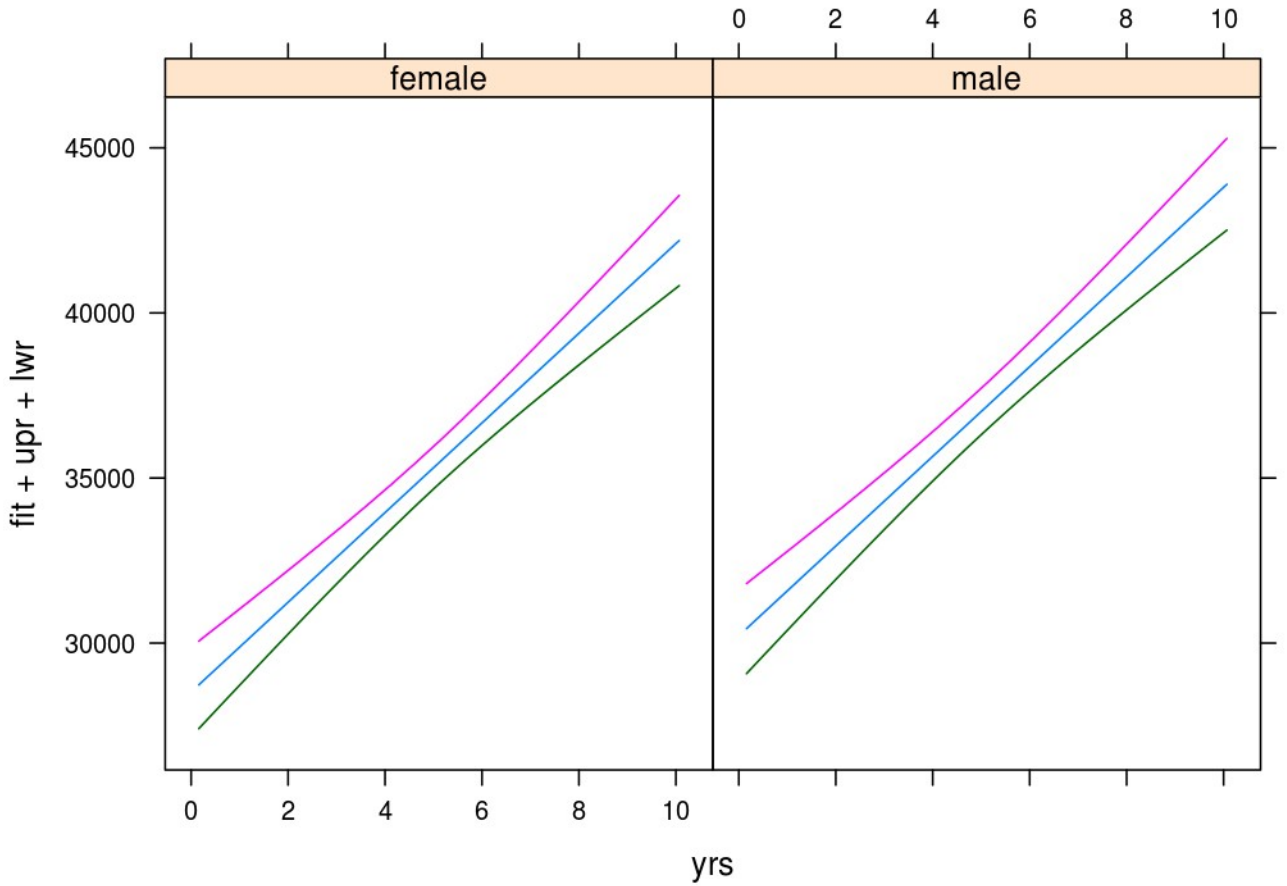
```
mod <- lm(salary.pre~visible+sex+edu+yrs*visible,
          data=ss)

pred <- expand.grid(visible=levels(ss$visible),
          sex=levels(ss$sex),
          edu=vals(ss$edu),
          yrs=vals(ss$yrs))
pred <- cbind(pred, predict(mod, newdata=pred,
          interval="confidence"))

vals <- function(x){a <- seq(from=-2.5, to=2.5,
by=.25)
       mean(x, na.rm=T)+a*sd(x, na.rm=T)
}
```

```
> head(pred)

visible    sex      edu       yrs      fit      lwr      upr
     no female 1.746594 0.1588892 18435.42 16731.68 20139.16
    yes female 1.746594 0.1588892 15851.93 13695.53 18008.33
     no   male 1.746594 0.1588892 20138.46 18325.99 21950.92
    yes   male 1.746594 0.1588892 17554.97 15276.26 19833.69
     no female 2.778601 0.1588892 19465.34 17818.71 21111.98
    yes female 2.778601 0.1588892 16881.86 14774.78 18988.93
```

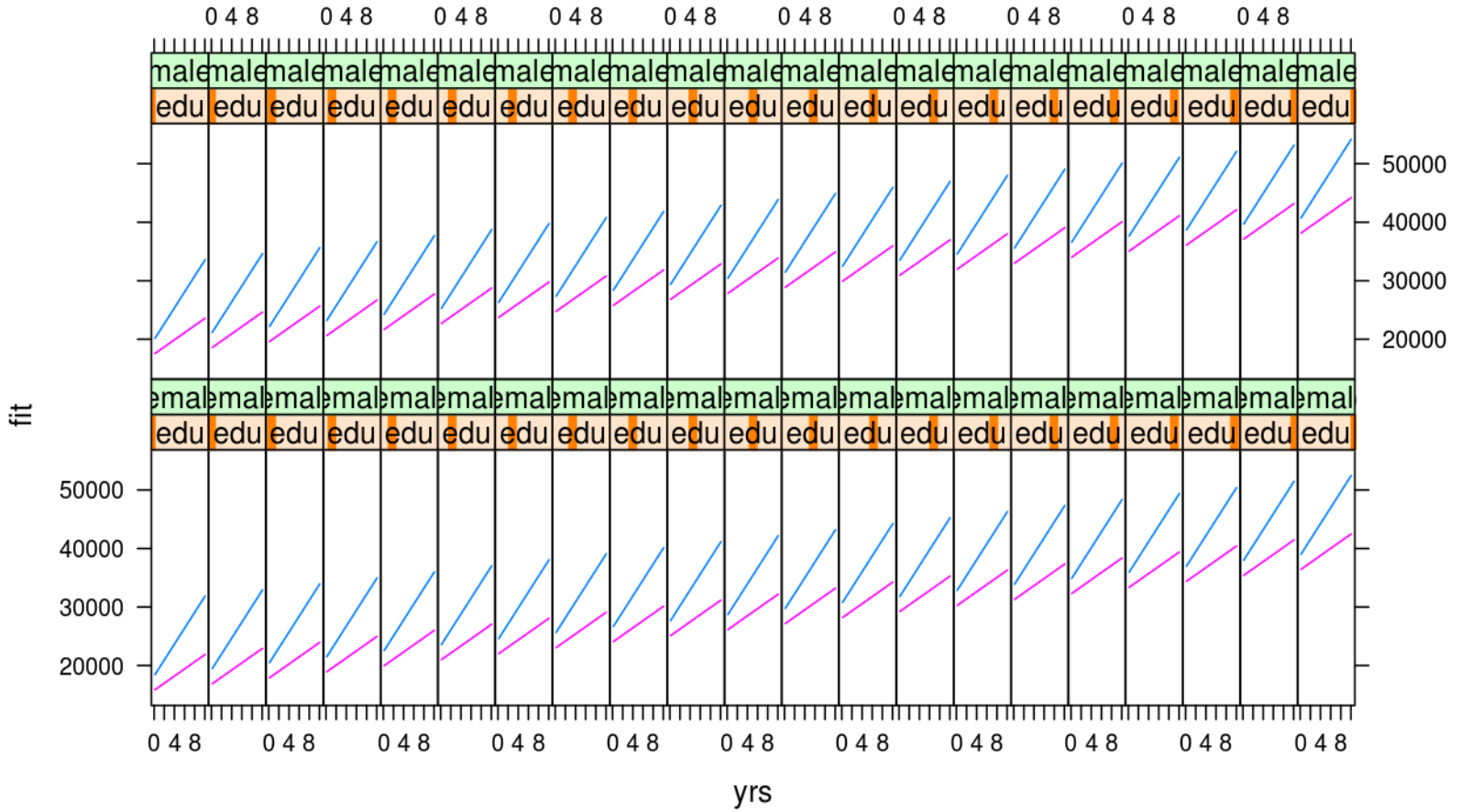
```
> xyplot(fit+upr+lwr ~ yrs | sex, data=pred, type="l")
```



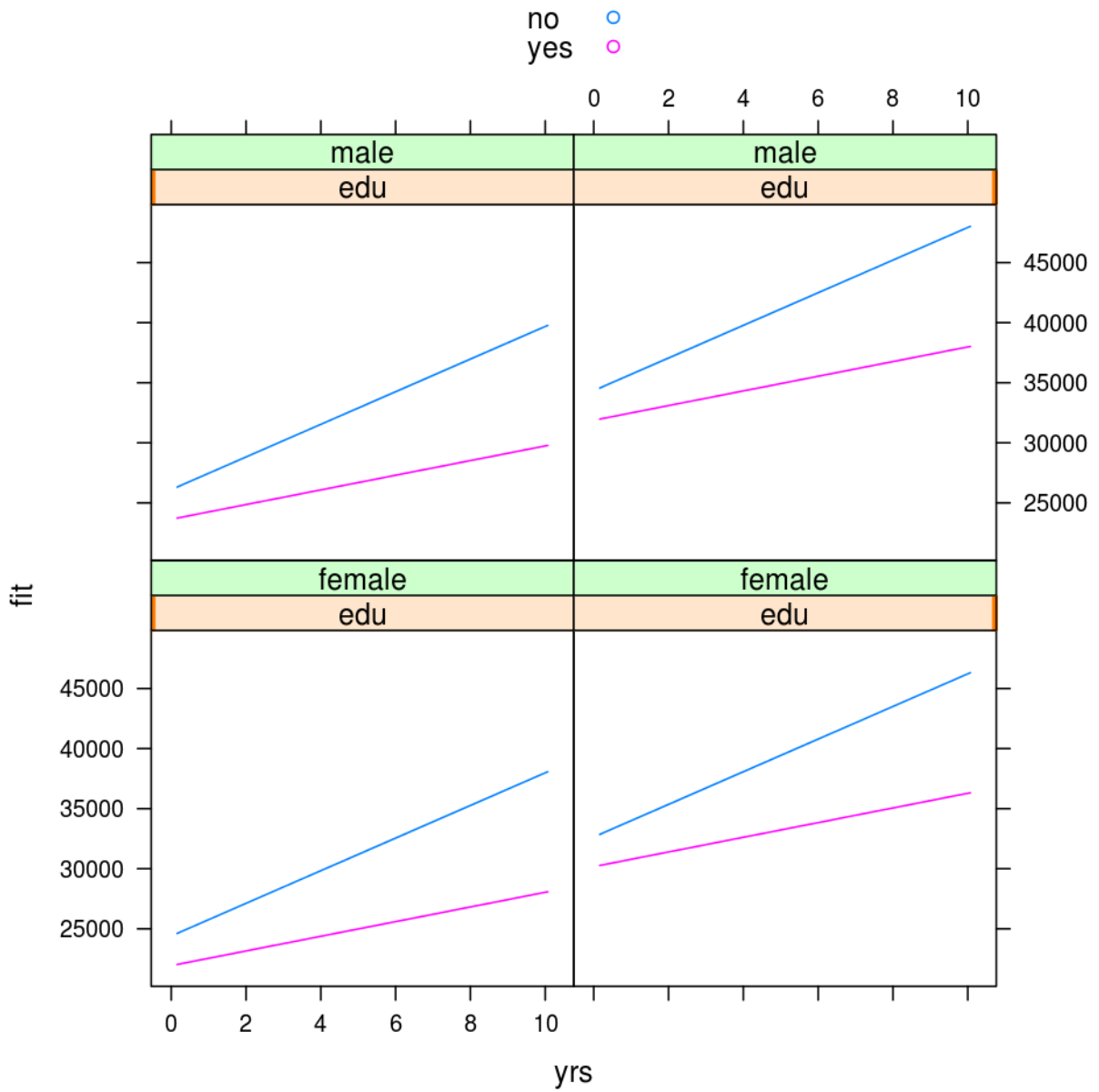
```
> xyplot(fit+upr+lwr ~ yrs | sex,
      data=subset(pred, edu==mean(ss$edu) & visible=="no"),
      type="l")
```



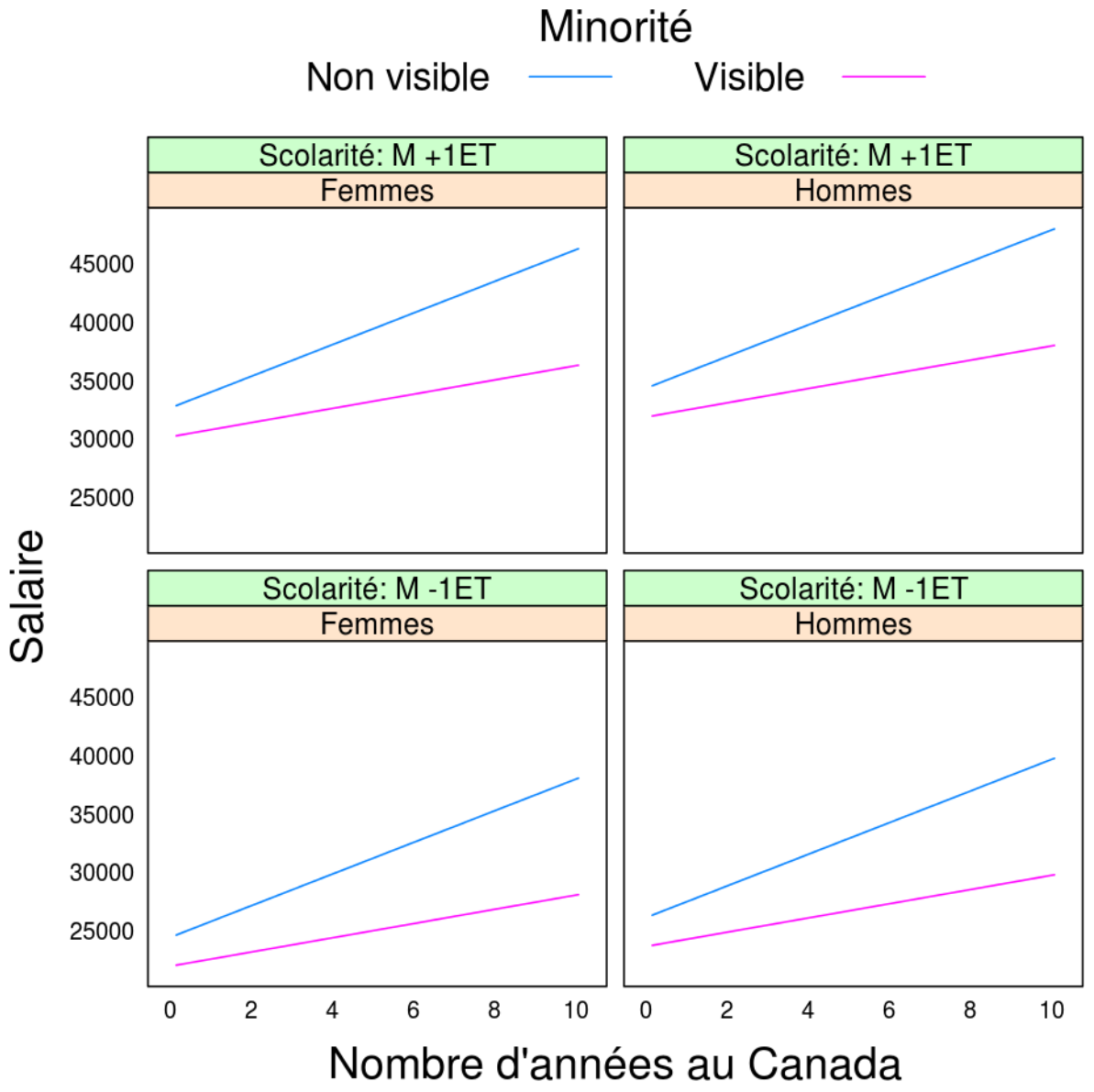
```
> xyplot(fit ~ yrs | edu+sex, groups=visible, data=pred,
         type="l")
```



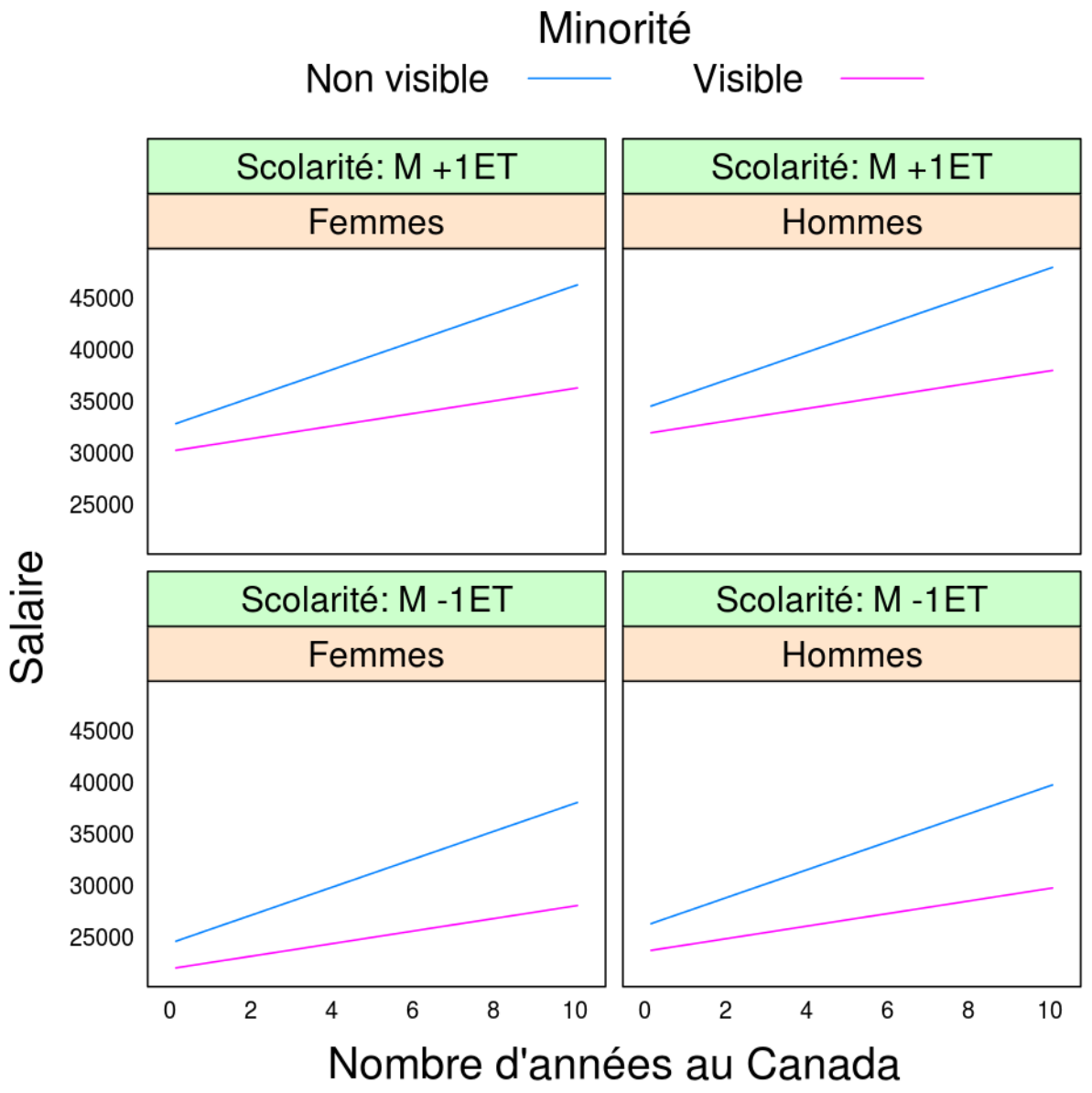
```
> xyplot(fit ~ yrs  | edu+ sex, groups=visible,
      data=subset(pred, edu %in% subvals(pred$edu, ss$edu, c(-1,1))),
      type="l")
```



```
> xyplot(fit ~ yrs  | factor(sex, labels=c("Femmes", "Hommes")) +
        factor(edu, labels=c("Scolarité: M -1ET",
              "Scolarité: M +1ET")),
        groups=visible,
        data=subset(pred, edu %in% subvals(pred$edu,
                    ss$edu, c(-1,1))),
        type="l",
        xlab=list("Nombre d'années au Canada", cex=1.5),
        ylab=list("Salaire", cex=1.5),
        layout=c(2,2),
        between=list(x=.5, y=.5),
        auto.key=list(title="Minorité", cex.title=1.5,
                      space="top", columns=2, points=F, lines=T,
                      text=c("Non visible", "Visible"), cex=1.3),
        scales=list(alternating=1, tck=0))
```

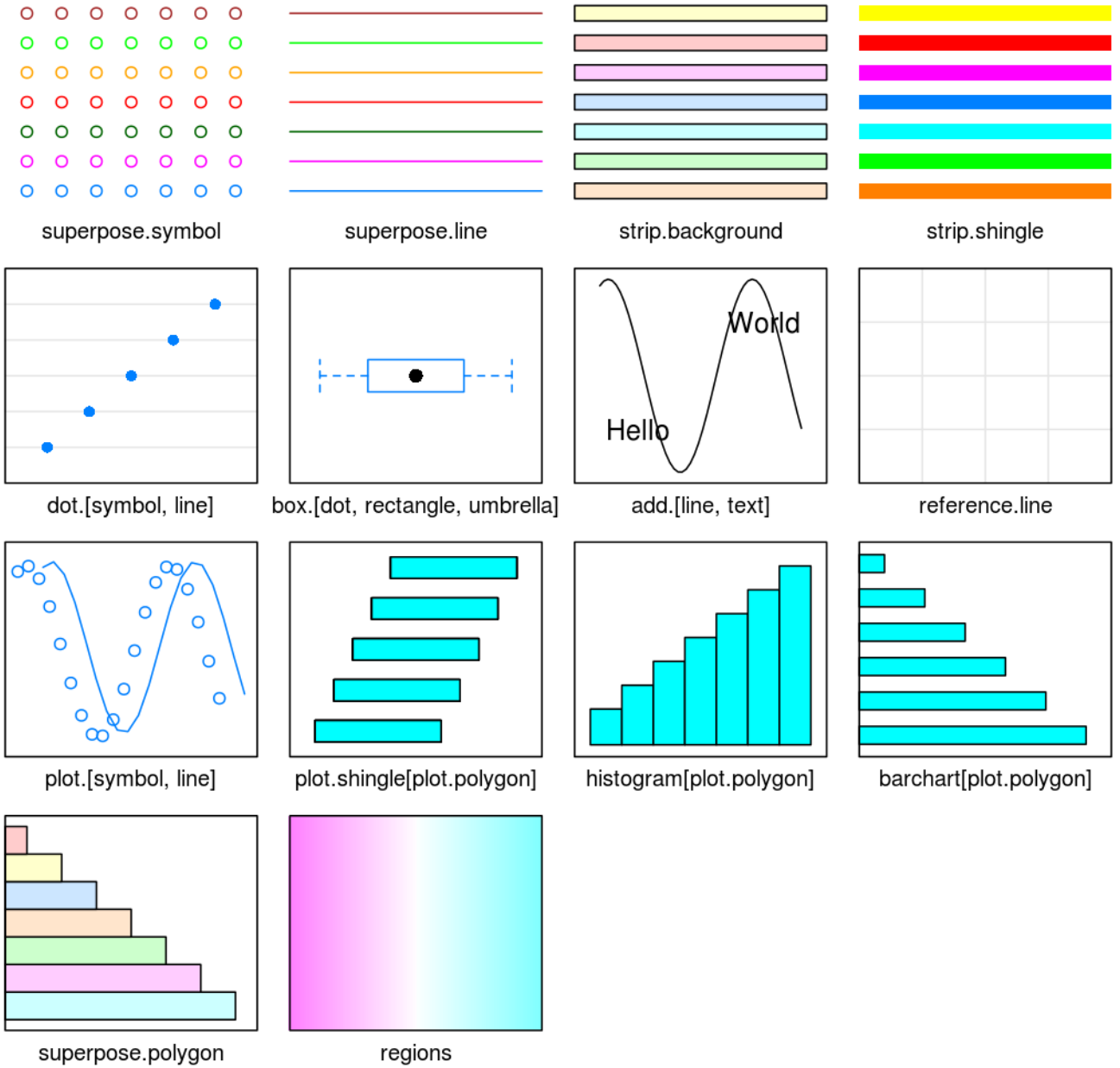

Minorité
Non visible
Visible
Scolarité: M +1ET
Femmes
Scolarité: M +1ET
Hommes
Scolarité: M -1ET
Femmes
Scolarité: M -1ET
Hommes
Salaire
45000
40000
35000
30000
25000
0
2
4
6
8
10
Nombre d'années au Canada

```
> ..., par.strip.text=list(cex=1.2, col="white", lines=1.3)
```



Couleurs :-(

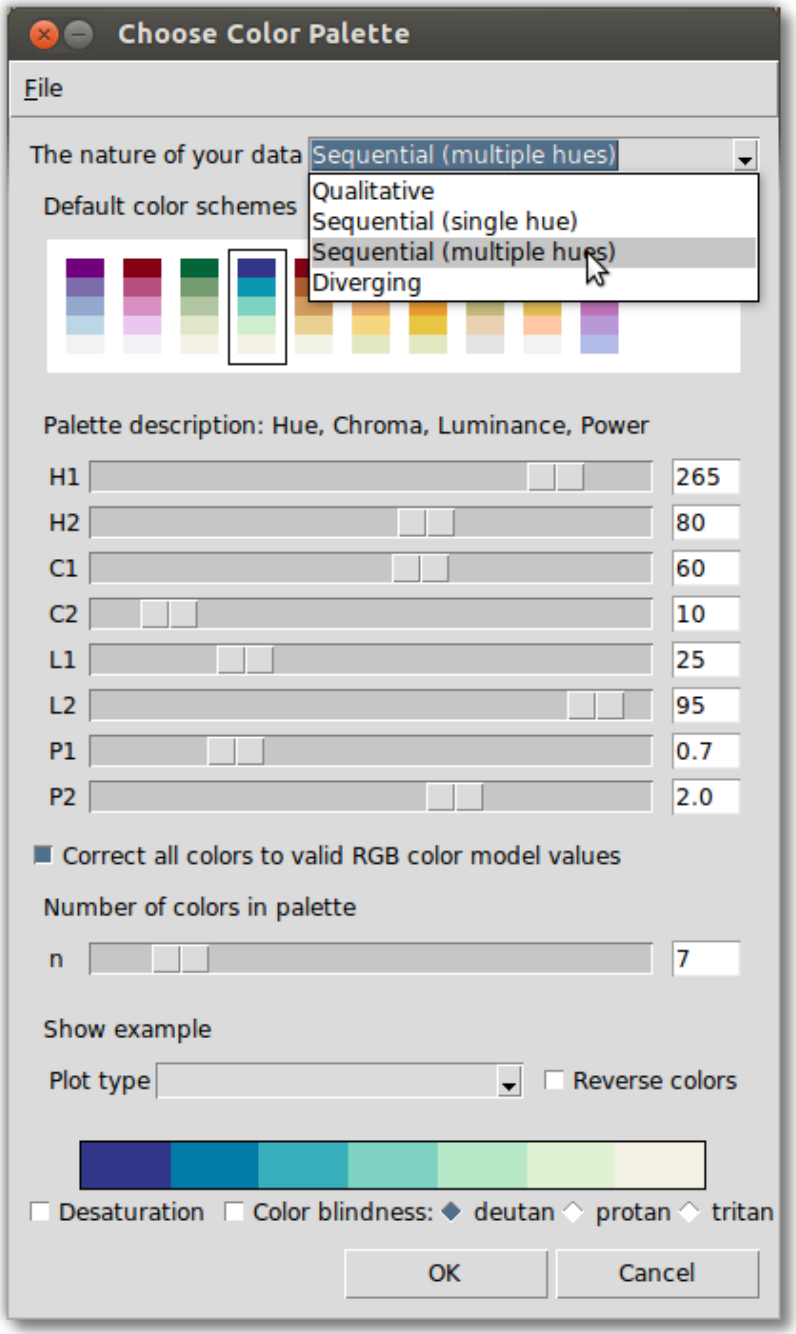
```
> show.settings()
```



Personnaliser:
1) Localement

```
> xyplot(fit ~ yrs  | factor(sex, labels=c("Femmes", "Hommes")) +
        factor(edu, labels=c("Scolarité: M -1ET",
              "Scolarité: M +1ET")),
        groups=visible,
        data=subset(pred, edu %in% subvals(pred$edu,
                    ss$edu, c(-1,1))),
        type="l",
        xlab=list("Nombre d'années au Canada", cex=1.5),
        ylab=list("Salaire", cex=1.5),
        layout=c(2,2),
        between=list(x=.5, y=.5),
        auto.key=list(title="Minorité", cex.title=1.5,
                      space="top", columns=2, points=F, lines=T,
                      text=c("Non visible", "Visible"), cex=1.3),
        scales=list(alternating=1, tck=0),
        par.strip.text=list(cex=1.2, col="white", lines=1.3),
        par.settings=list(...))
```

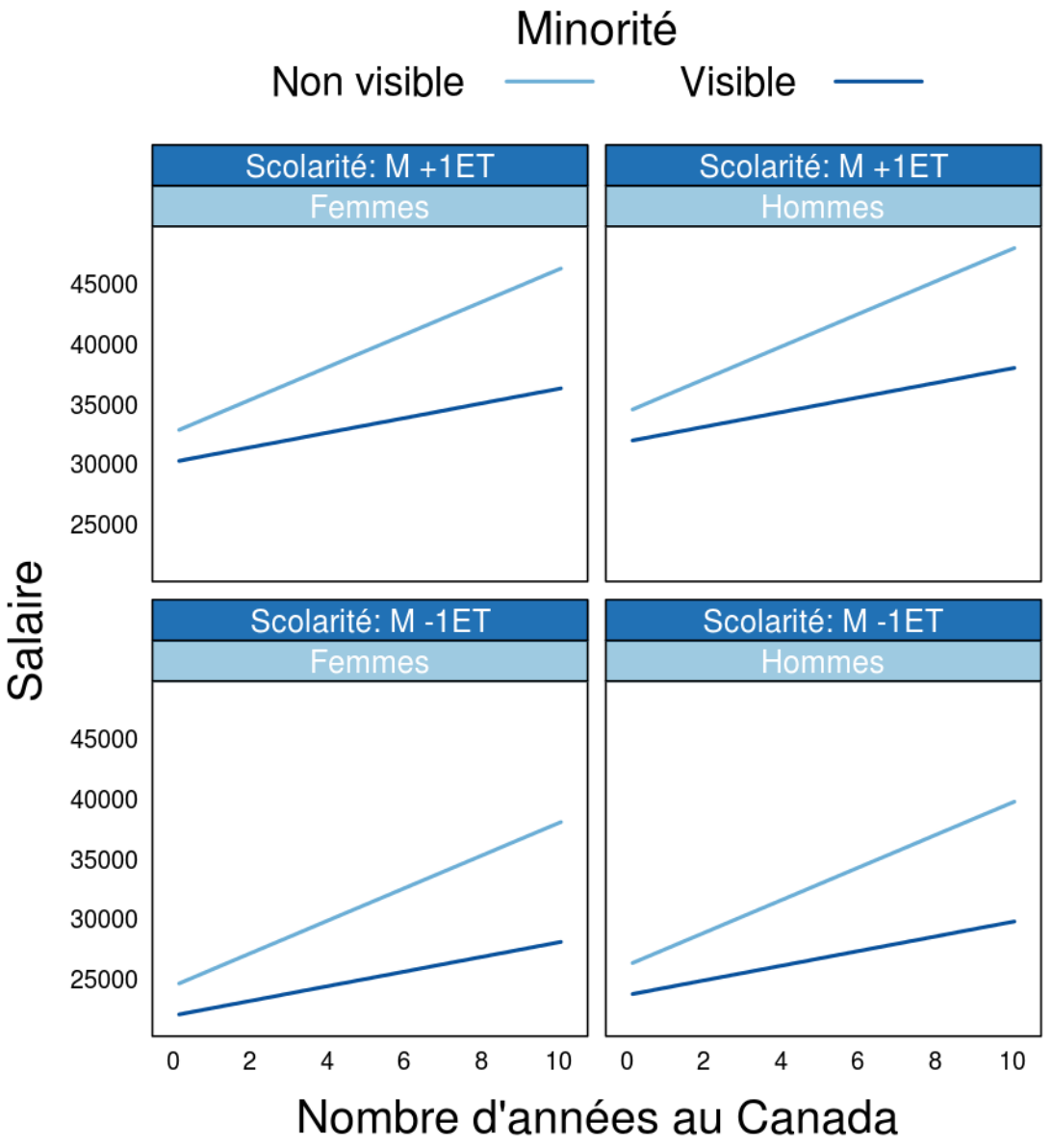
```
> library(colorspace)
> choose_palette()
```



```
> library(RColorBrewer)
> display.brewer.all()
```

```
> mescouleurs <- brewer.pal(9,"Blues")

> mesoptions <- list(
    superpose.line=list(col=mescouleurs[c(5,8)],
                        lwd=2),
    strip.background=list(col=mescouleurs[c(4,7)]),
    strip.border=list(col="black"))
```

```
> xyplot(..., par.settings=mesoptions))
```



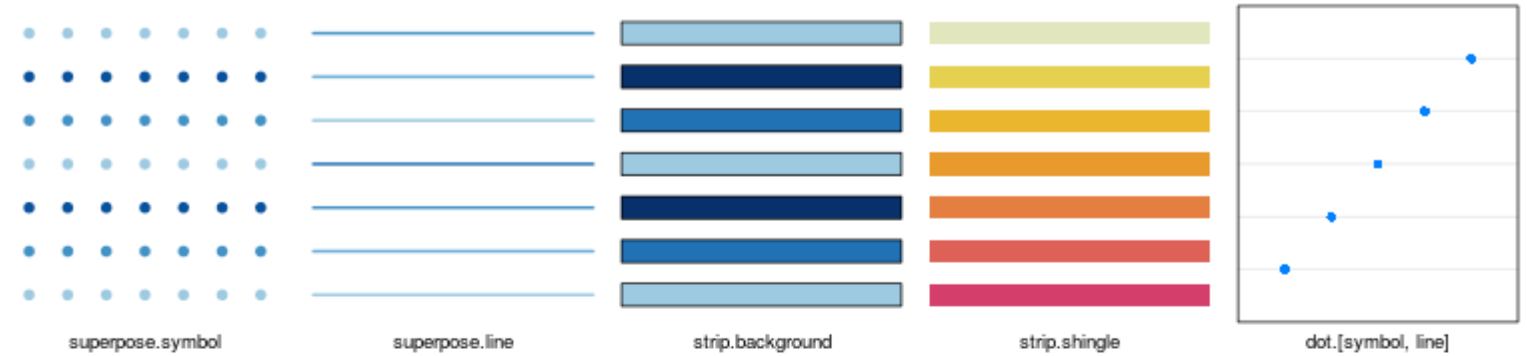
Personnaliser:
2) Globalement

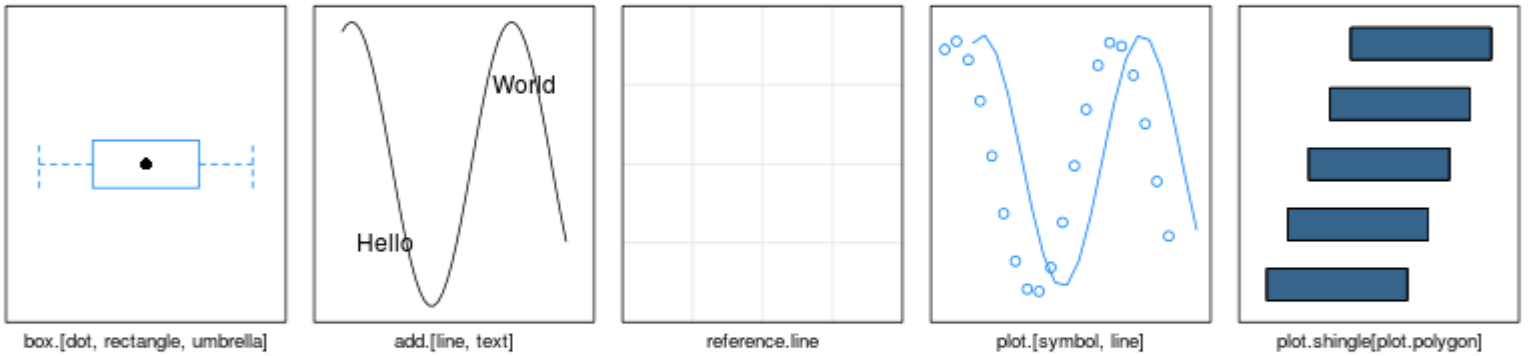
```
> trellis.par.get()
```

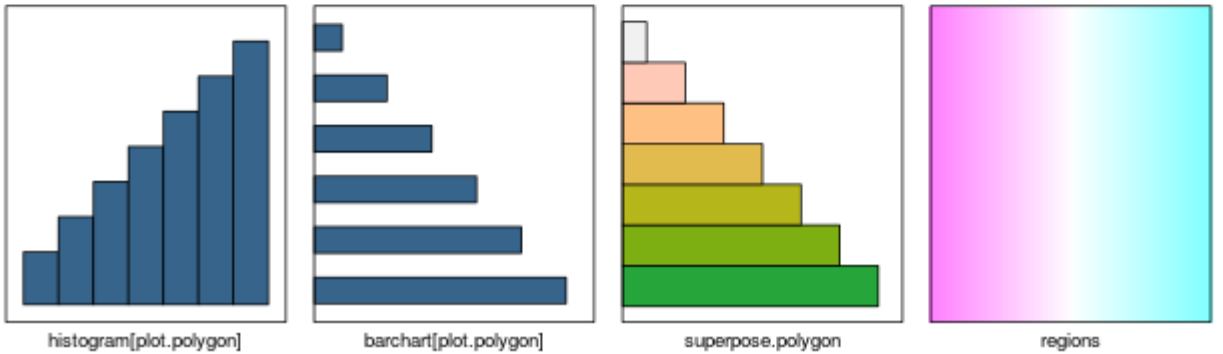
```
$grid.pars
list()
$fontsize
$fontsize$text
[1] 12
$fontsize$points
[1] 8
$background
$background$alpha
[1] 1
$background$col
[1] "transparent"
$panel.background
$panel.background$col
[1] "transparent"
$clip
$clip$panel
[1] "on"
$clip$strip
[1] "on"
$add.line
$add.line$alpha
[1] 1
$add.line$col
[1] "#000000"
$add.line$lty
[1] 1
$add.line$lwd
[1] 1
$add.text
$add.text$alpha
[1] 1
$add.text$cex
[1] 1
$add.text$col
[1] "#000000"
$add.text$font
[1] 1
$add.text$lineheight
[1] 1.2
$plot.polygon
$plot.polygon$alpha
[1] 1
$plot.polygon$col
[1] "#00ffff"
$plot.polygon$border
[1] "black"
$plot.polygon$lty
[1] 1
$plot.polygon$lwd
[1] 1
$box.dot
$box.dot$alpha
[1] 1
$box.dot$col
[1] "#000000"
$box.dot$cex
[1] 1
$box.dot$font
[1] 1
$box.dot$pch
[1] 16
$box.rectangle
$box.rectangle$alpha
[1] 1
$box.rectangle$col
[1] "#0080ff"
$box.rectangle$fill
[1] "transparent"
$box.rectangle$lty
[1] 1
$box.rectangle$lwd
[1] 1
$box.umbrella
$box.umbrella$alpha
[1] 1
$box.umbrella$col
[1] "#0080ff"
$box.umbrella$lty
```

etc.

```
> default <- trellis.par.get()

> montheme <- trellis.par.get()

> montheme$superpose.line$col <- brewer.pal(9,"Blues")[4:7]
> montheme$superpose.line$lwd <- 2
> montheme$superpose.polygon$col <- terrain_hcl(7)
> montheme$strip.background$col <- brewer.pal(9, "Blues")[c(4,7,9)]
> montheme$superpose.symbol$col <- brewer.pal(9, "Blues")[c(4,6,8)]
> montheme$superpose.symbol$pch <- 19
> montheme$strip.shingle$col <- heat_hcl(7)
> montheme$plot.polygon$col <- "steelblue4"

> trellis.par.set(montheme)

> show.settings()
```


superpose.symbol
superpose.line
strip.background
strip.shingle
dot.[symbol, line]


World
Hello
box.[dot, rectangle, umbrella]
add.[line, text]
reference.line
plot.[symbol, line]
plot.shingle[plot.polygon]


histogram[plot.polygon]
barchart[plot.polygon]
superpose.polygon
regions

ggplot2

(Hadley Wickham)

Données

Années	Salaire	Sexe
7	46000	Homme
6	35000	Femme
5	40000	Homme
7	37000	Femme
4	32000	Homme
5	38000	Femme

Mappage

Années	Salaire	Sexe
7	46000	Homme
6	35000	Femme
5	40000	Homme
7	37000	Femme
4	32000	Homme
5	38000	Femme

x	y	couleur
7	46000	Homme
6	35000	Femme
5	40000	Homme
7	37000	Femme
4	32000	Homme
5	38000	Femme

Échelle

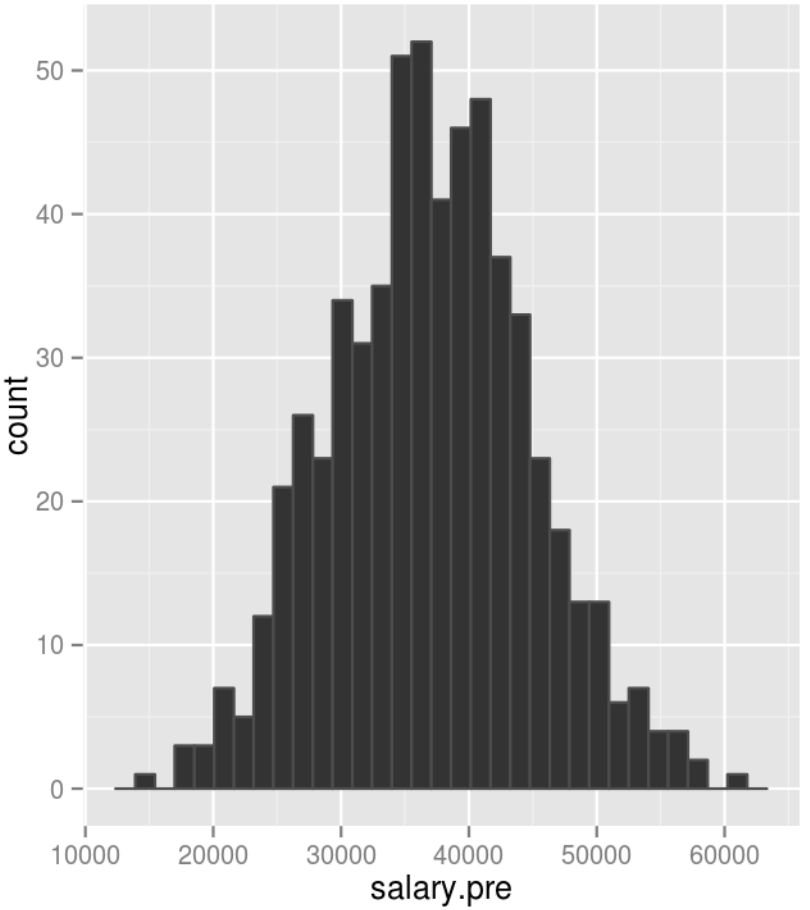
Années	Salaire	Sexe
7	46000	Homme
6	35000	Femme
5	40000	Homme
7	37000	Femme
4	32000	Homme
5	38000	Femme

x	y	couleur
.96	.94	#00C1A9
.68	.24	#FF6C91
.36	.54	#00C1A9
.96	.36	#FF6C91
.04	.06	#00C1A9
.36	.42	#FF6C91

Composantes

- Données
- Attributs esthétiques
- Objets géometriques
- Transformation statistique
- Échelles
- Systèmes de coordonnées
- Guides
- Autres paramètres

- geom = barres
- stat = bin
- échelle = linéaire
- coord = cartésien
- guide = quadrillage



```
ggplot(data, mapping) +
layer(
      stat="",
      geom="",
      geom_params=list(),
      stat_param=list()
      ) +
scale() + theme() + annotations() + ...
```

Données et mappage

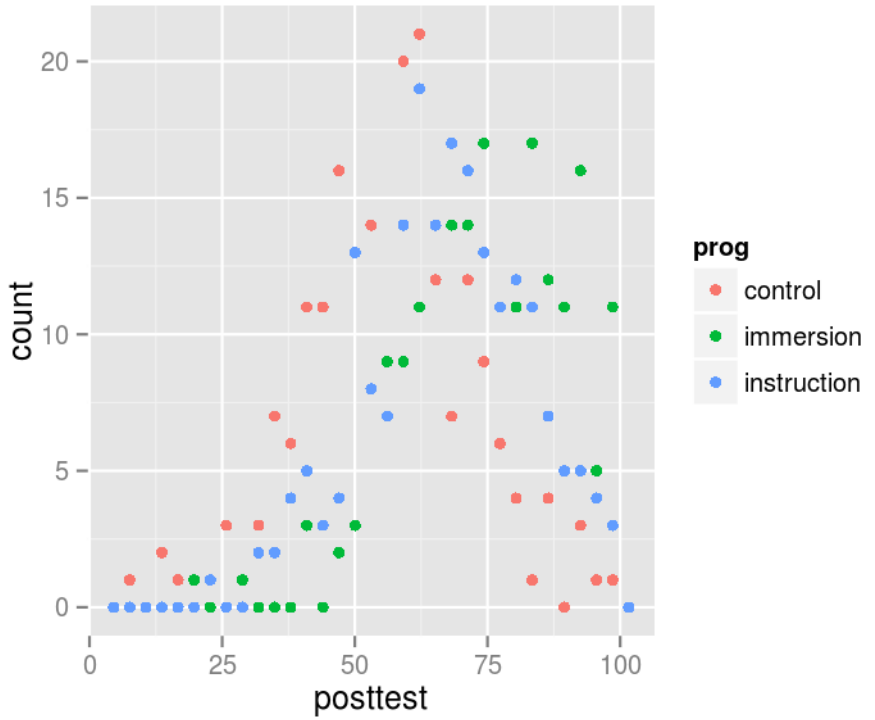
```
> p <- ggplot(data=ss,
                aes(x=posttest, fill=prog, colour=prog))
```

geom & stat

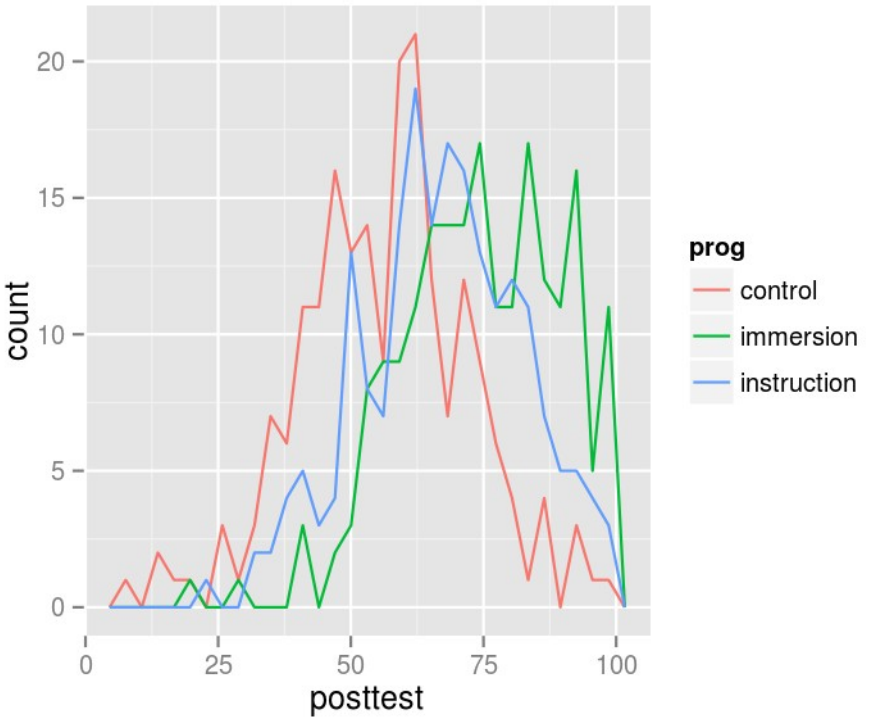
- Séparation de la transformation des données de leur représentation graphique

geom & stat

```
> p + geom_point(stat="bin")
```




```
> p + geom_line(stat="bin")
```

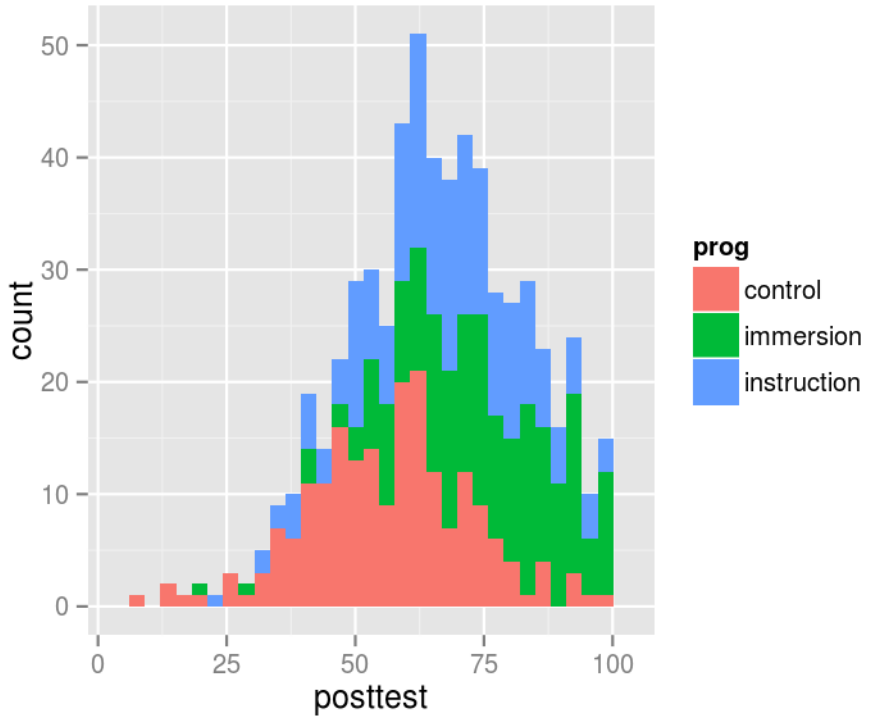


geom & stat

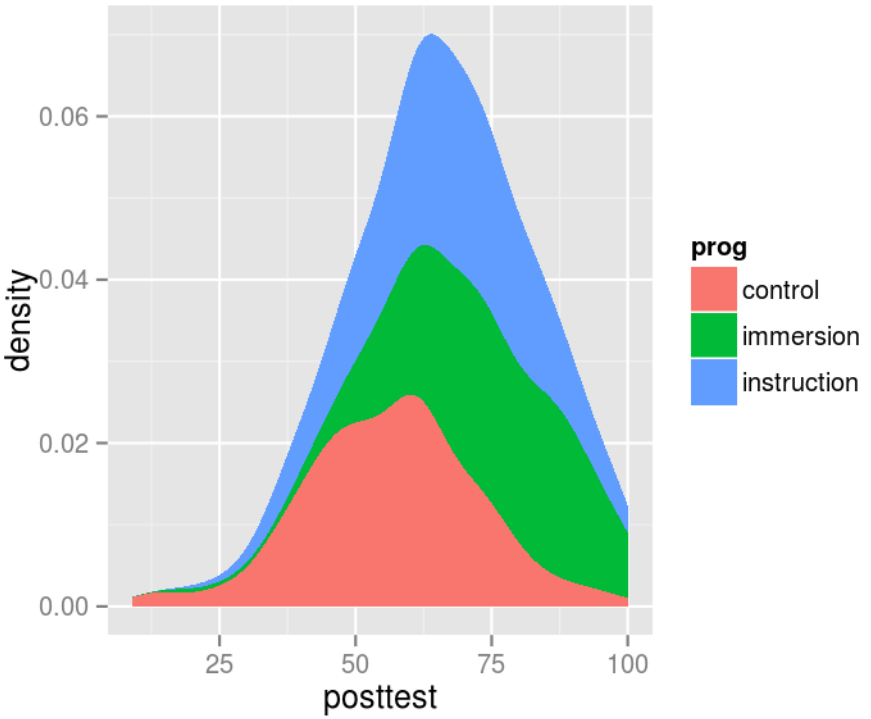
- Défauts respectifs

geom & stat

```
> p + geom_bar()
```



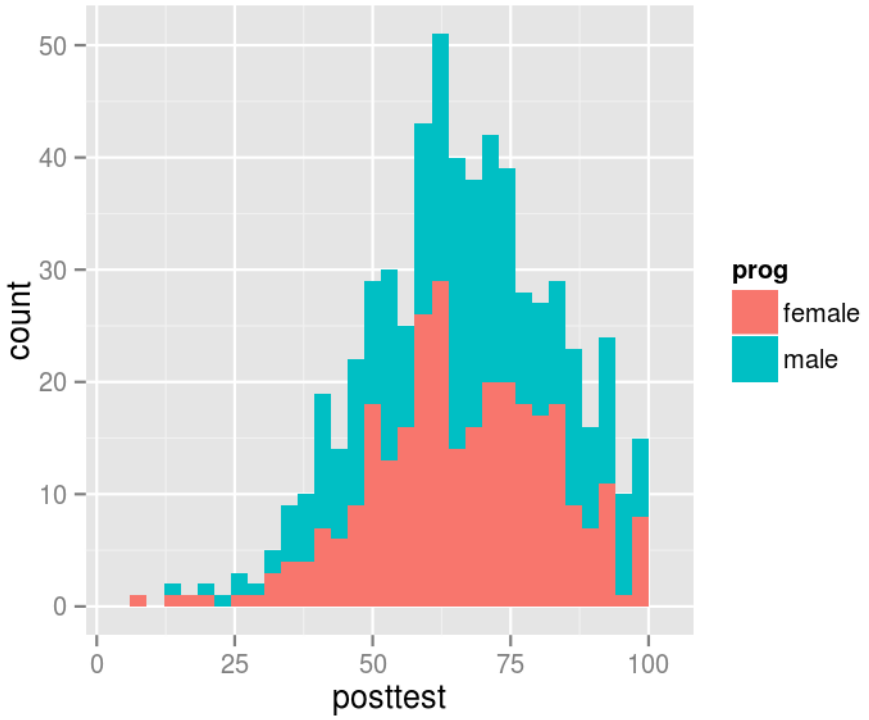

```
> p + stat_density()
```



geom & stat

- Ajouter/modifier des mappages 'données-attributs esthétiques'

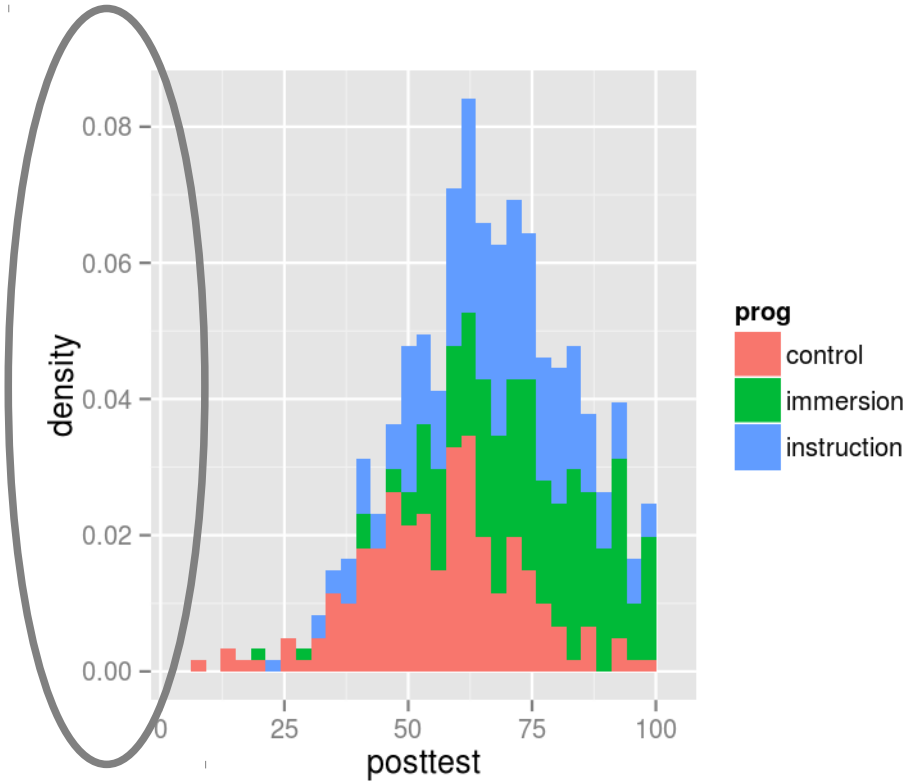
```
> p+geom_bar(aes(fill=sex, colour=NULL))
```



geom & stat

- Variables calculées par ggplot

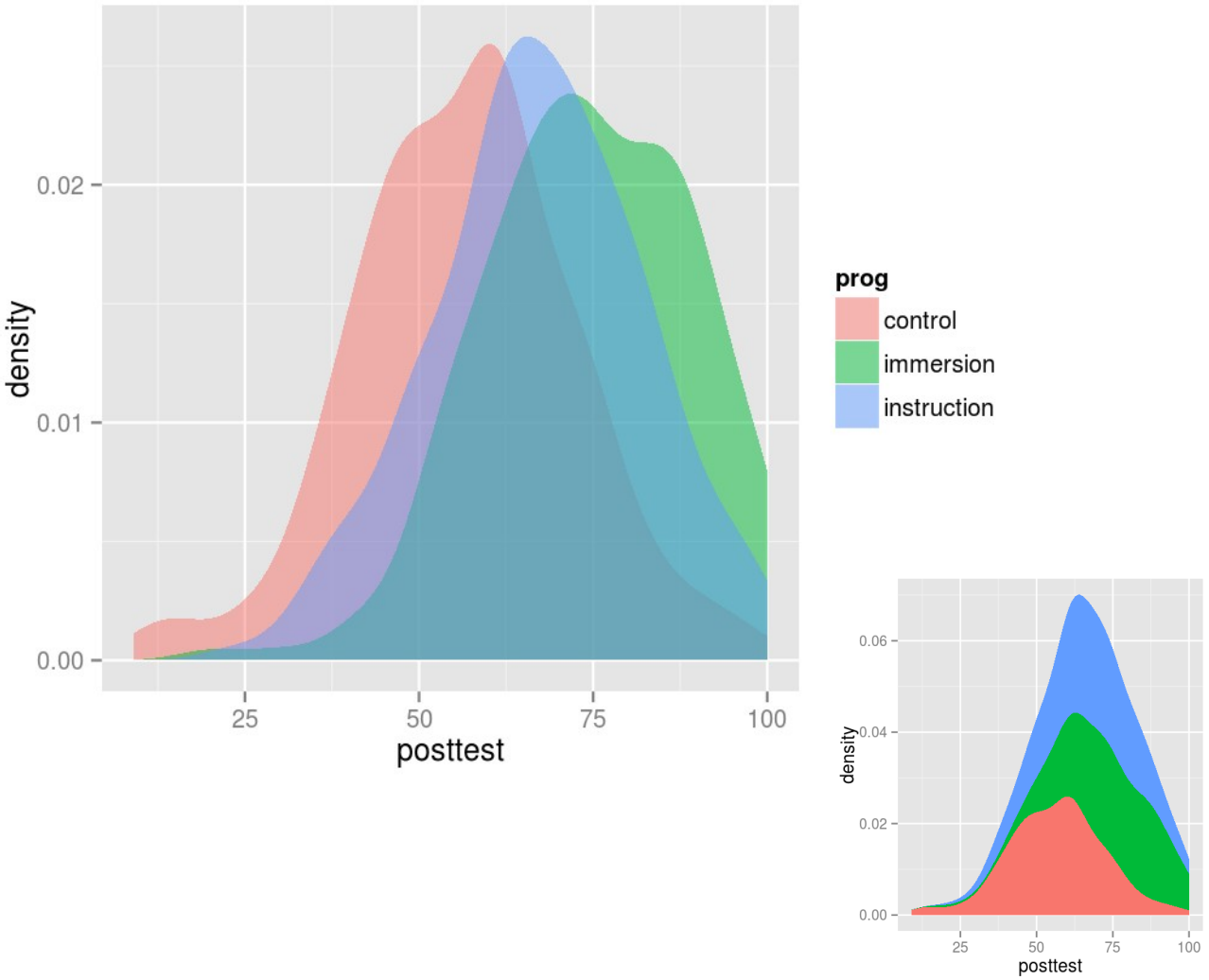
```
> p+geom_bar(aes(y=..density..))
```



geom & stat

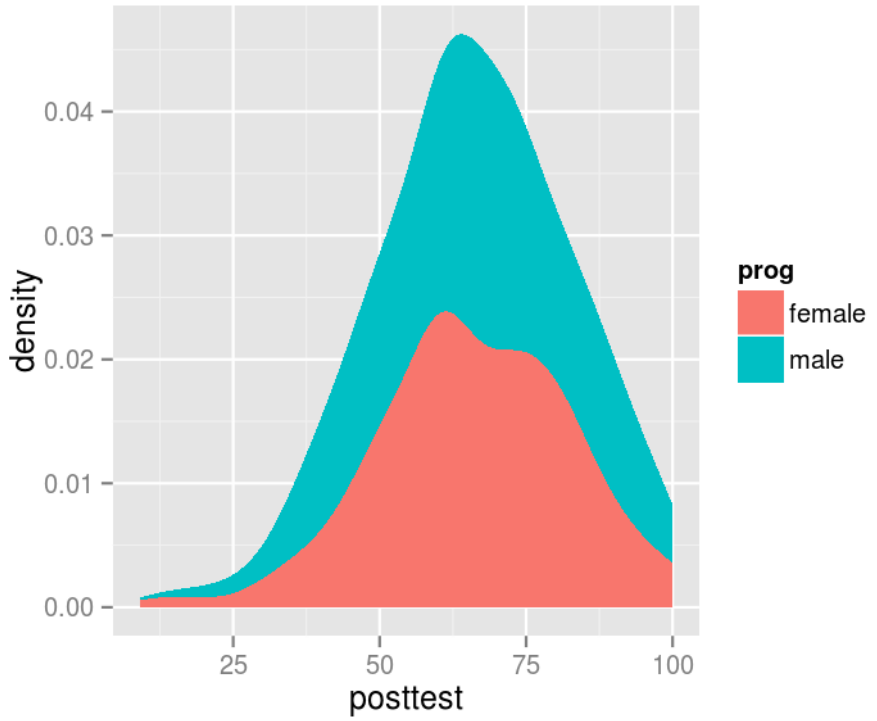
- Paramètres supplémentaires

```
> +stat_density(position="identity", alpha=.5)
```

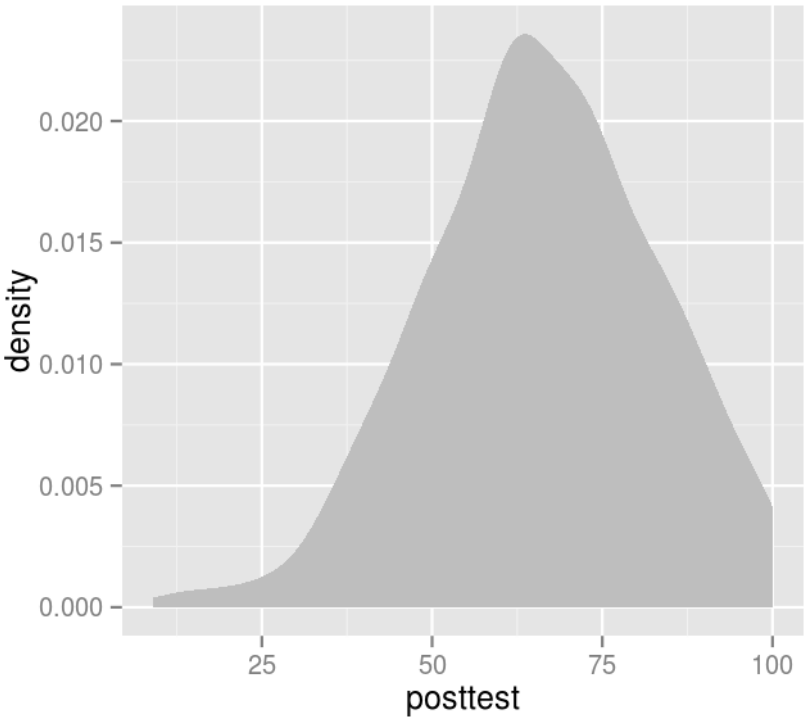


Mappage vs. Réglage

```
> +geom_area(stat="density",
             aes(fill=sex))
```



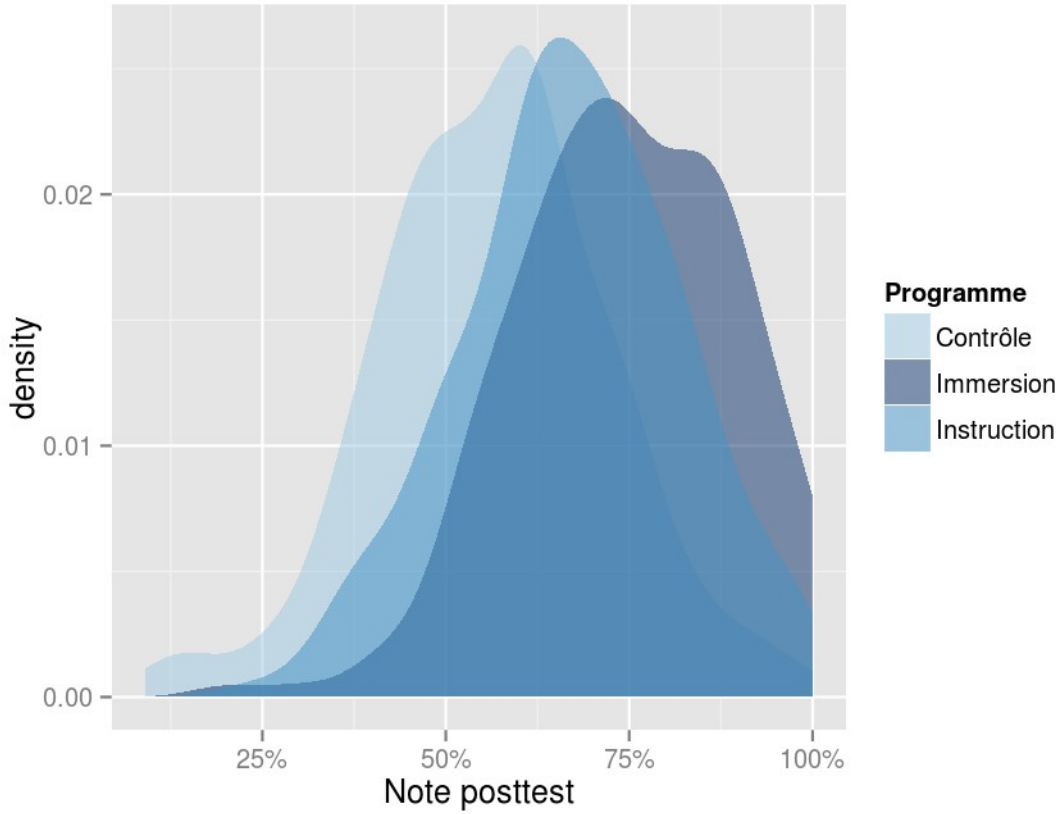

```
> +geom_area(stat="density",
             fill="grey")
```



scale

- Contrôle mappage + guide visuel

```
> p+stat_density(position="identity", alpha=.5, aes(colour=NULL)) +
   scale_fill_manual(name="Programme",
                     breaks=c("control","immersion","instruction"),
                     labels=c("Contrôle","Immersion","Instruction"),
                     values=mescouleurs[c(4,9,6)]) +
   scale_x_continuous(name="Note posttest",
                      breaks=c(25, 50, 75, 100),
                      labels=c("25%", "50%", "75%", "100%"))
```

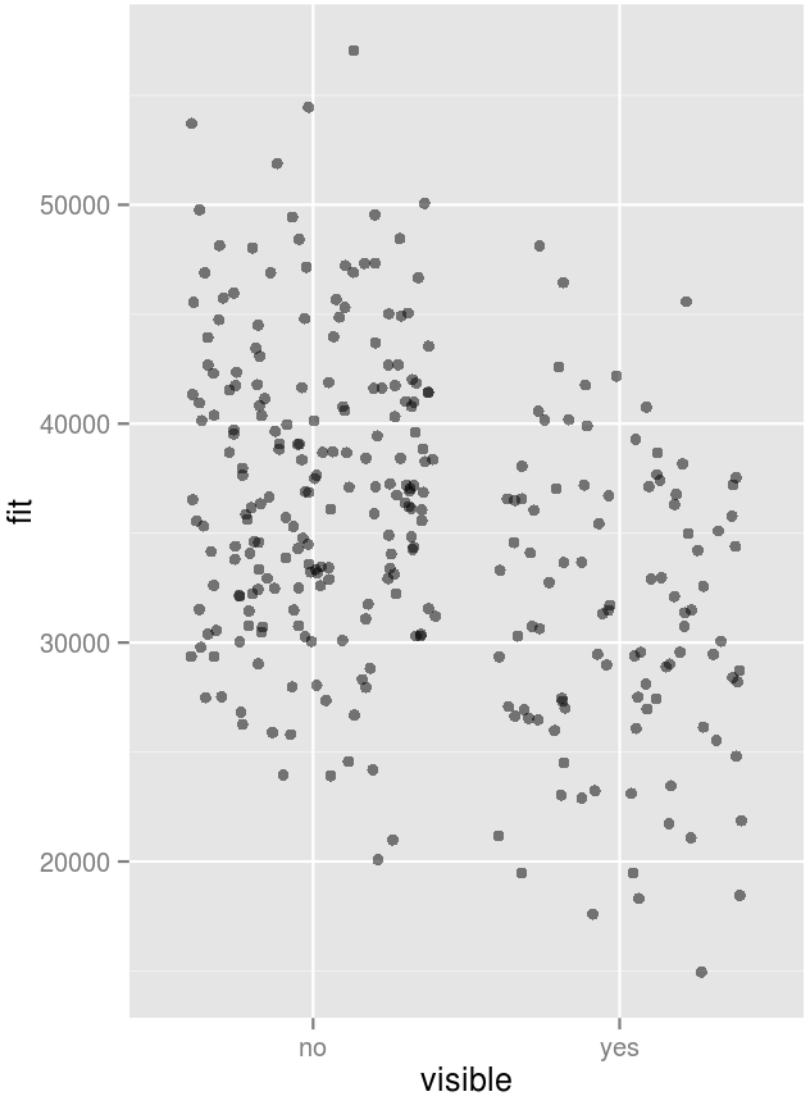


Couche après couche

```
> q <- ggplot(data=pred, aes(x=visible, y=fit))
```

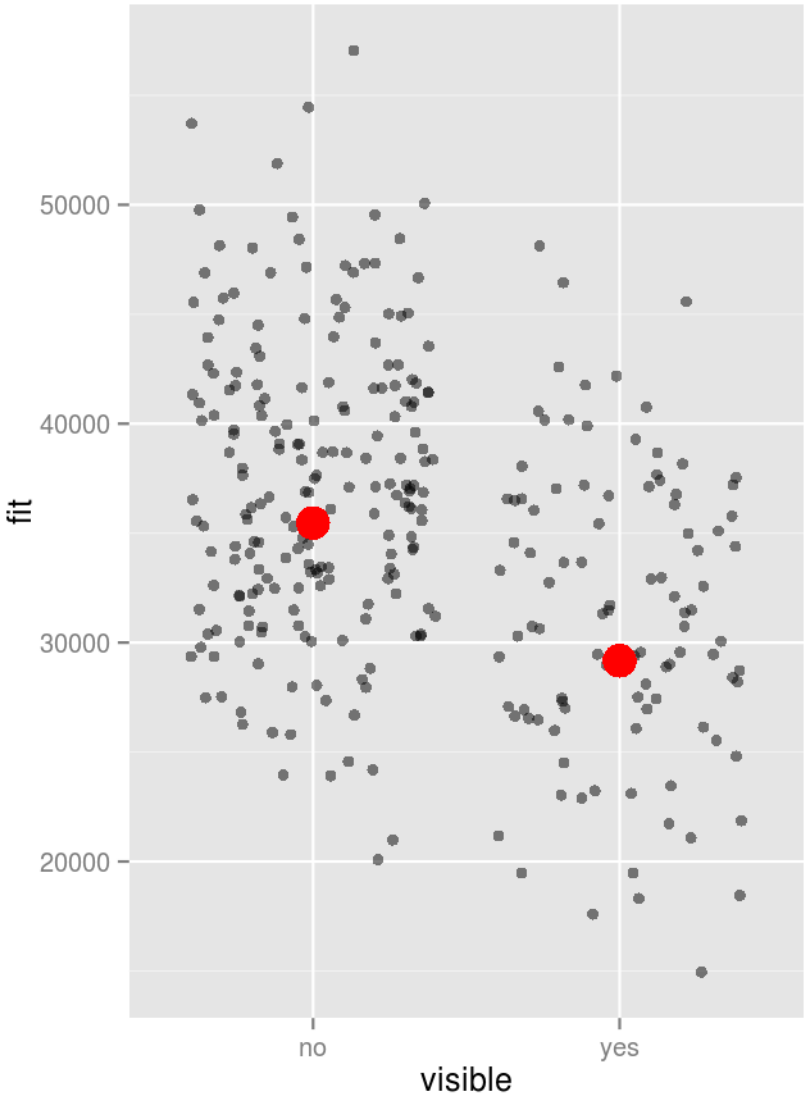
Couche après couche

```
+ geom_jitter(data=ss,
               aes(x=visible, y=salary.pre),
               alpha=.5)
```



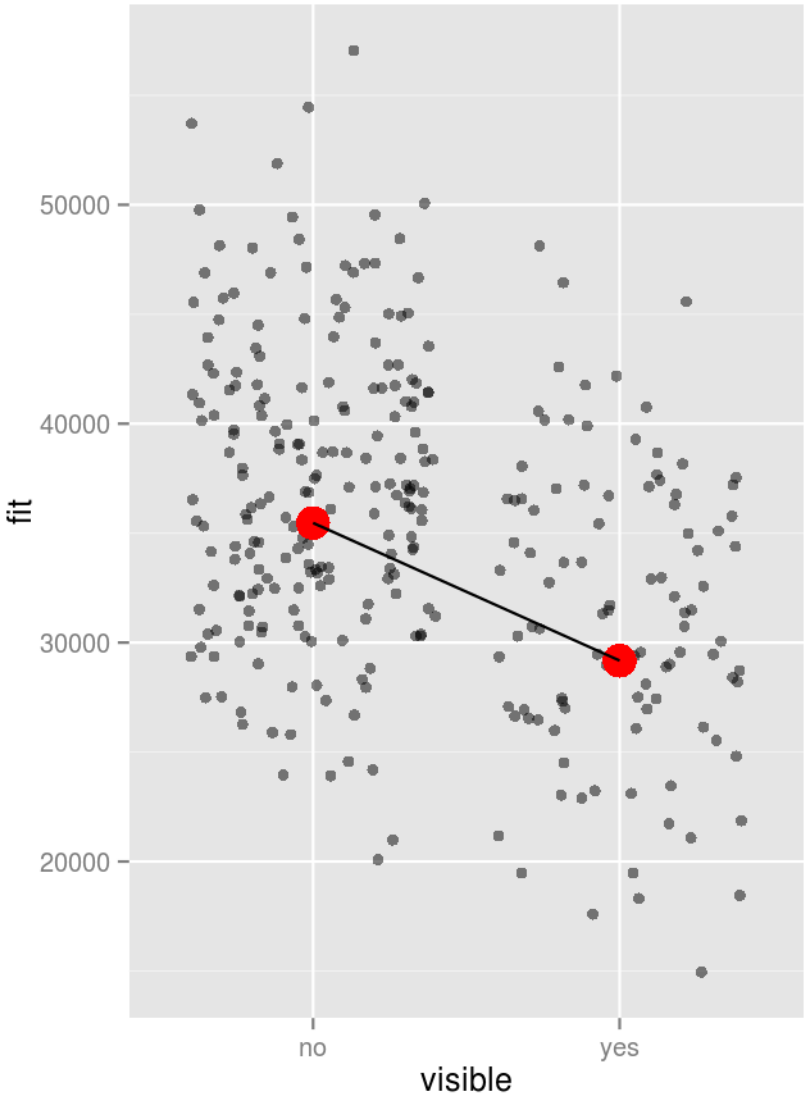
Couche après couche

```
+ geom_point(colour="red", size=6)
```



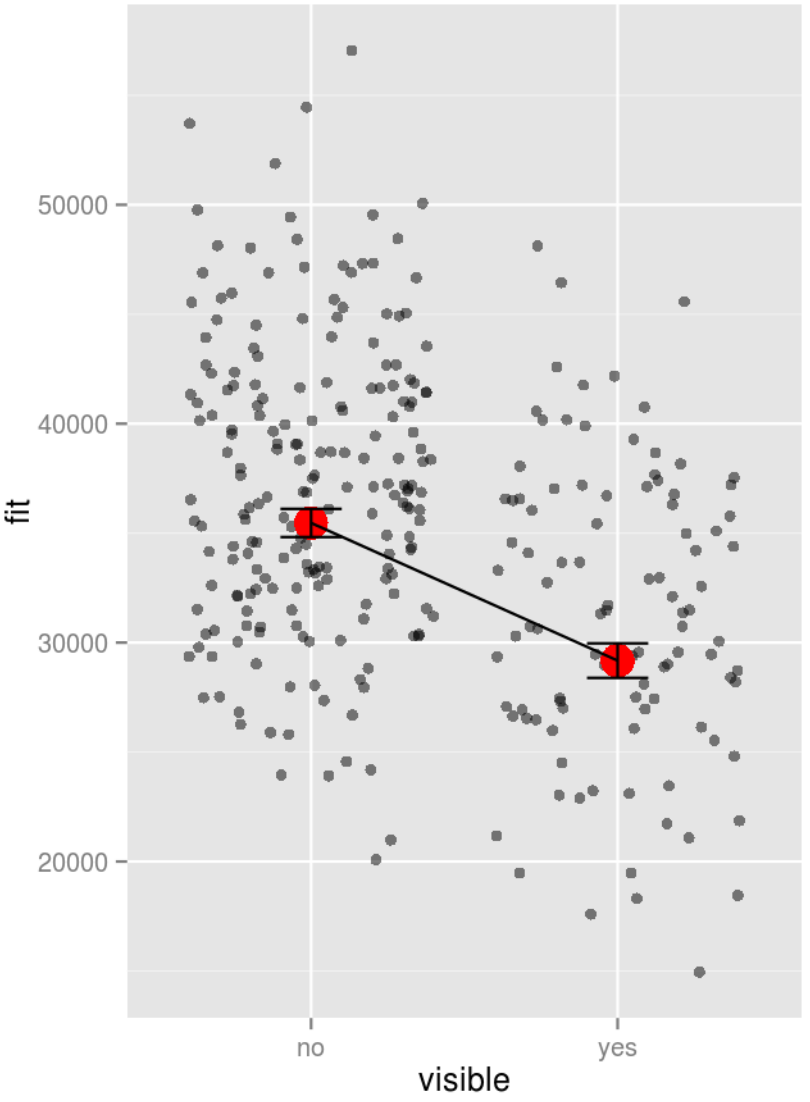
Couche après couche

```
+ geom_line(aes(group=1))
```



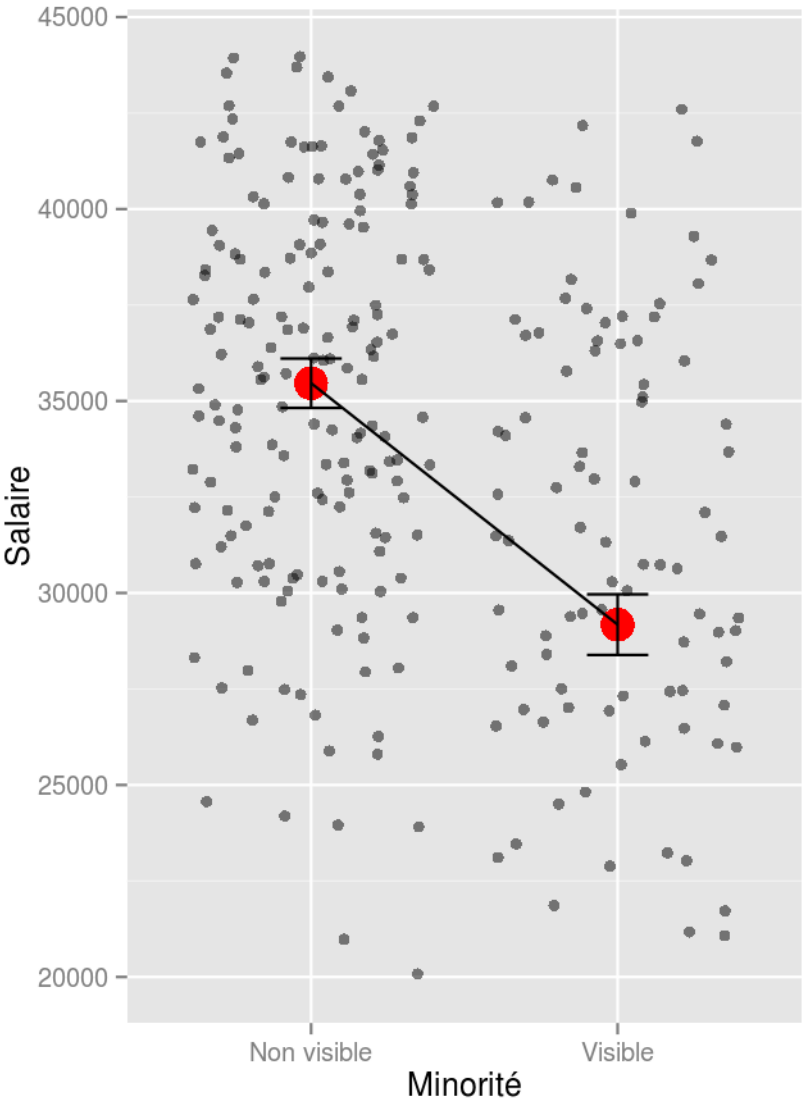
Couche après couche

```
+ geom_errorbar(aes(ymin=lwr,
                    ymax=upr),
                width=.2)
```



Couche après couche

```
+ scale_y_continuous(name="Salaire",
                      limits=c(20000,44000))
+ scale_x_discrete(name="Minorité",
                    breaks=c("no", "yes"),
                    labels=c("Non visible",
                             "Visible")
```

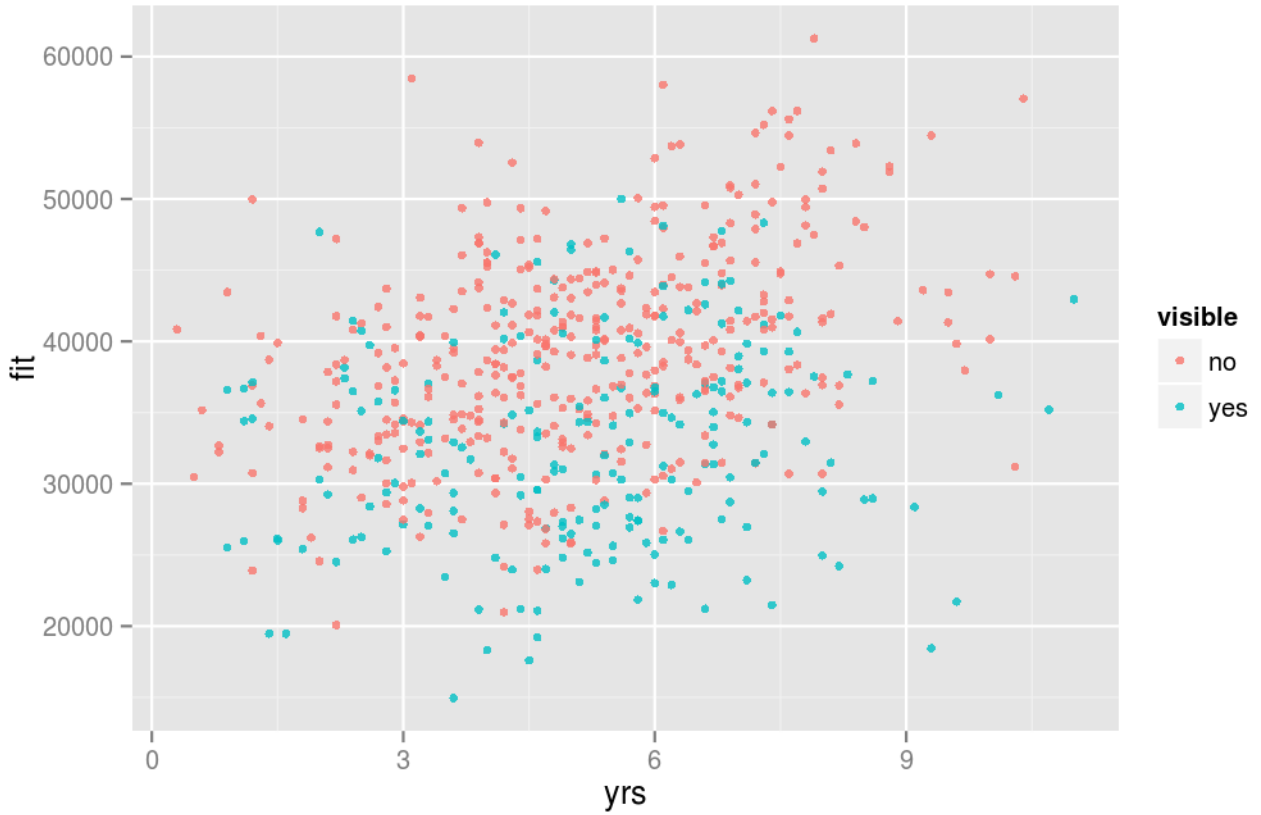


group

```
> r <- ggplot(data=pred, aes(x=yrs, y=fit))
```

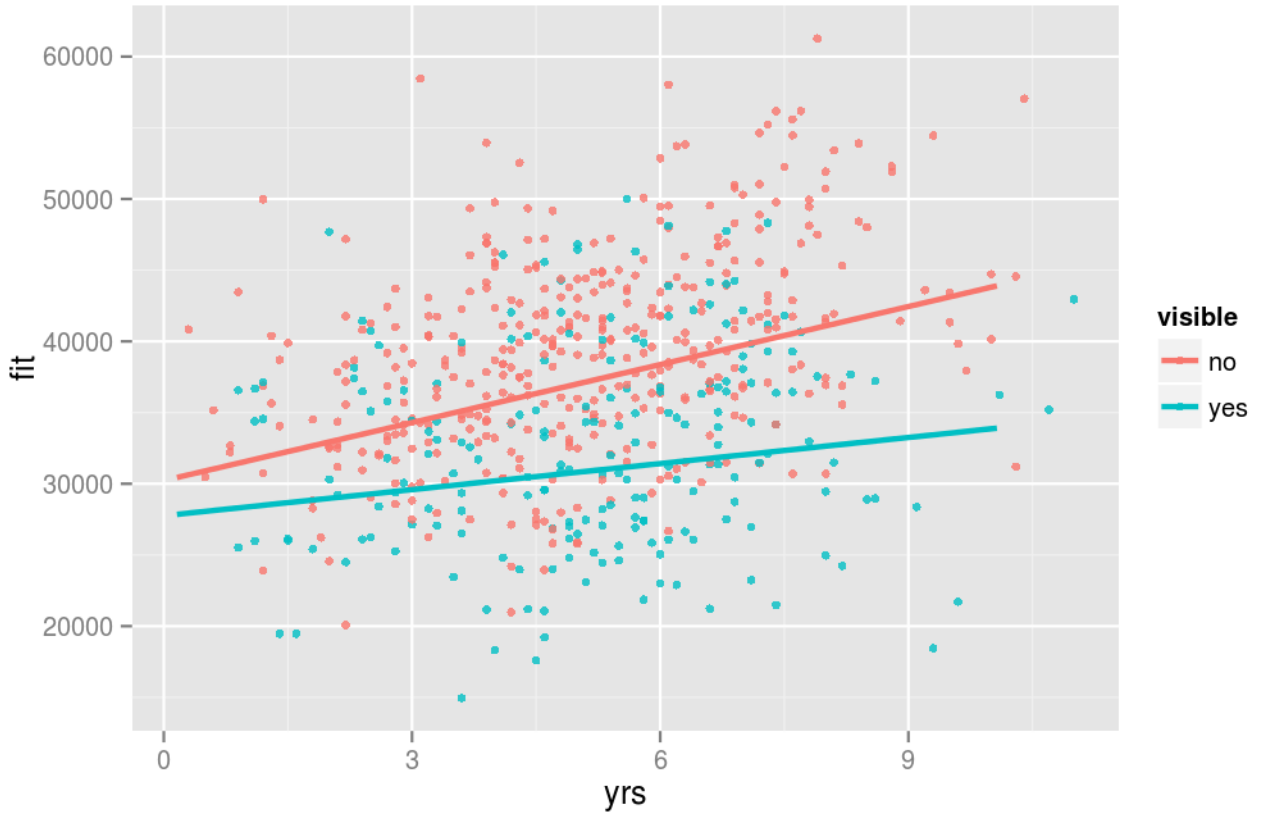
group

```
> r + geom_point(data=ss, aes(y=salary.pre, colour=visible),
                 size=1.5, alpha=.8)
```



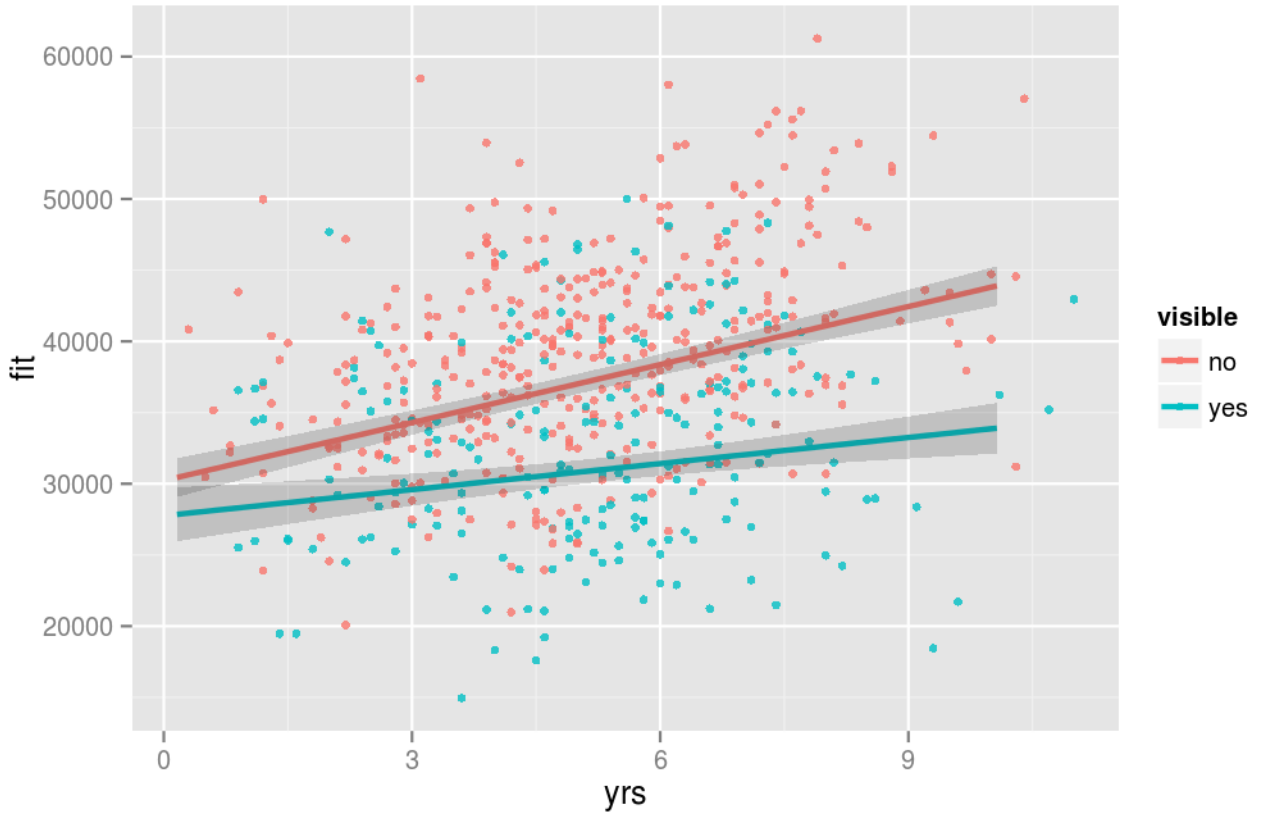
group

```
> + geom_line(aes(group=visible, colour=visible), size=1)
```



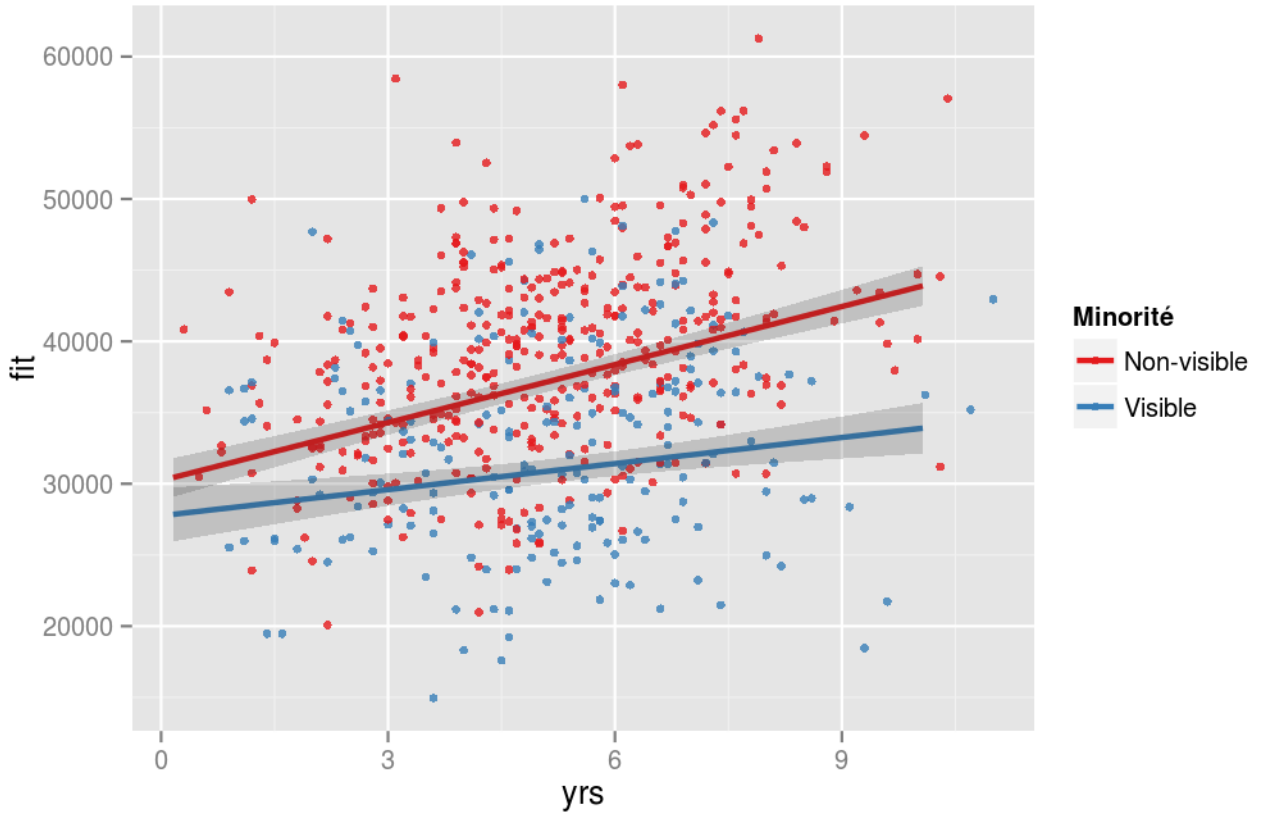
group

```
> + geom_ribbon(aes(ymin=lwr, ymax=upr, group=visible), alpha=.2)
```



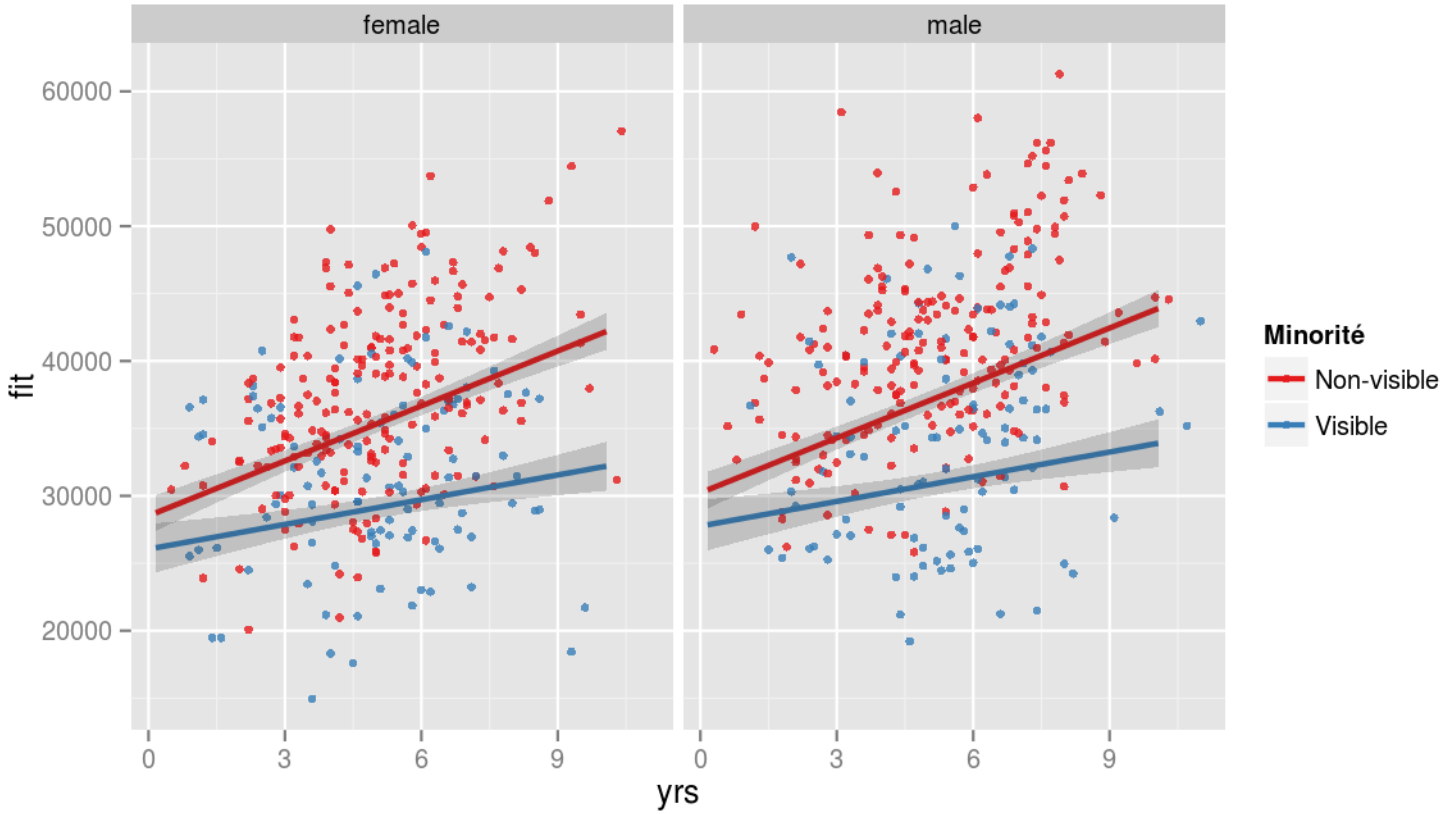
group

```
> + scale_colour_brewer(name="Minorité", breaks=c("no", "yes"),
                        labels=c("Non-visible", "Visible"),
                        palette="Set1")
```



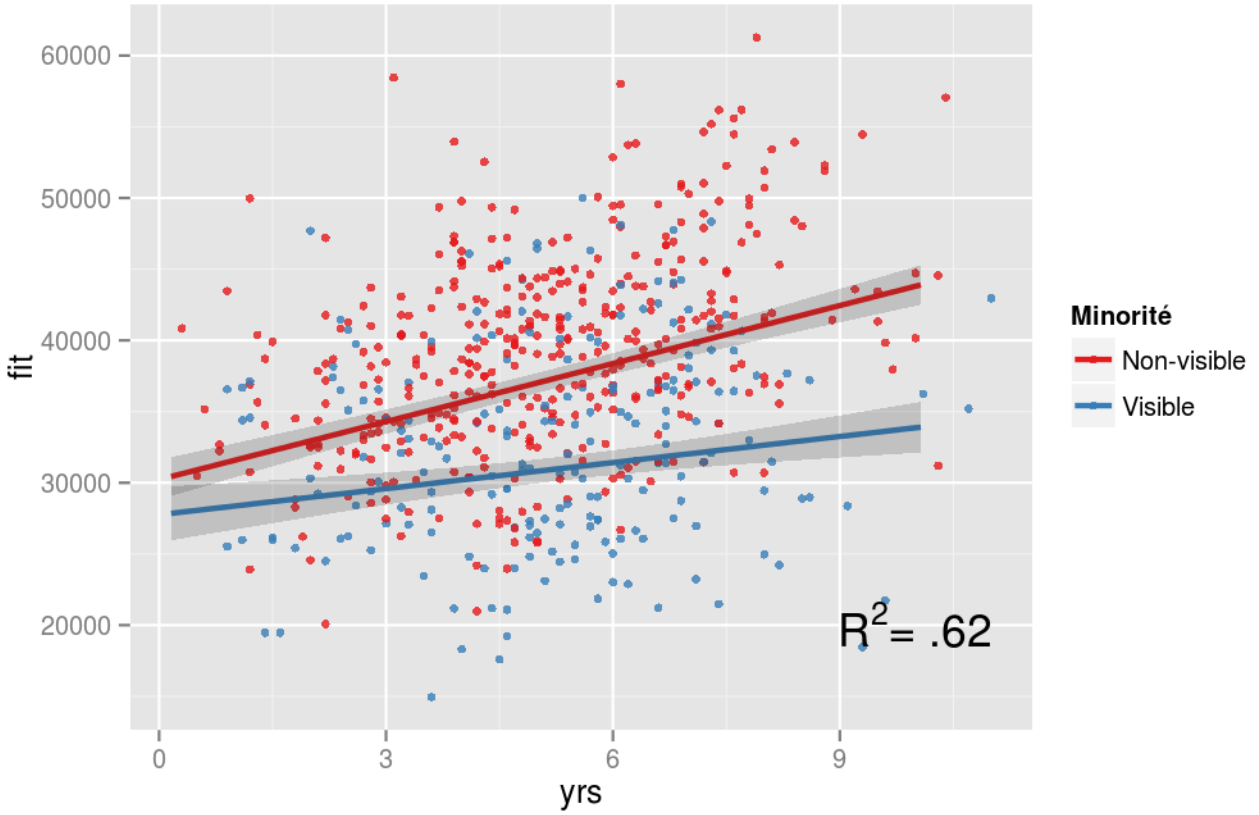
facet

```
> + facet_grid(.~sex)
```



annotate

```
> + annotate("text", x=10, y=20000,
                label="R^2 * '= .62'", size=6, parse=T)
```



theme

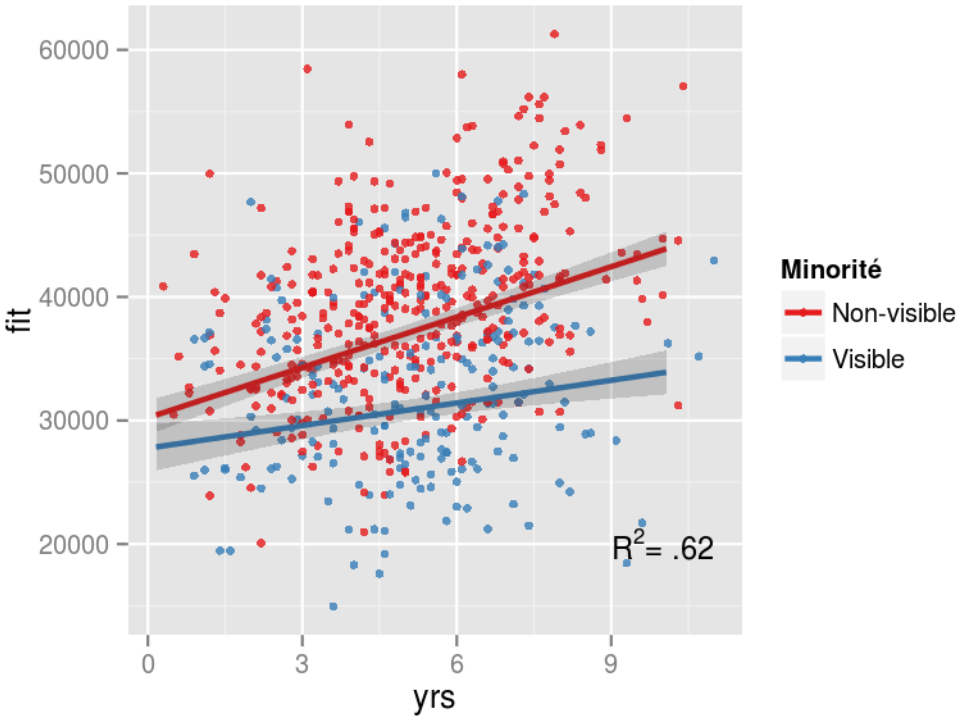
- Séparation données vs. Non-données

theme

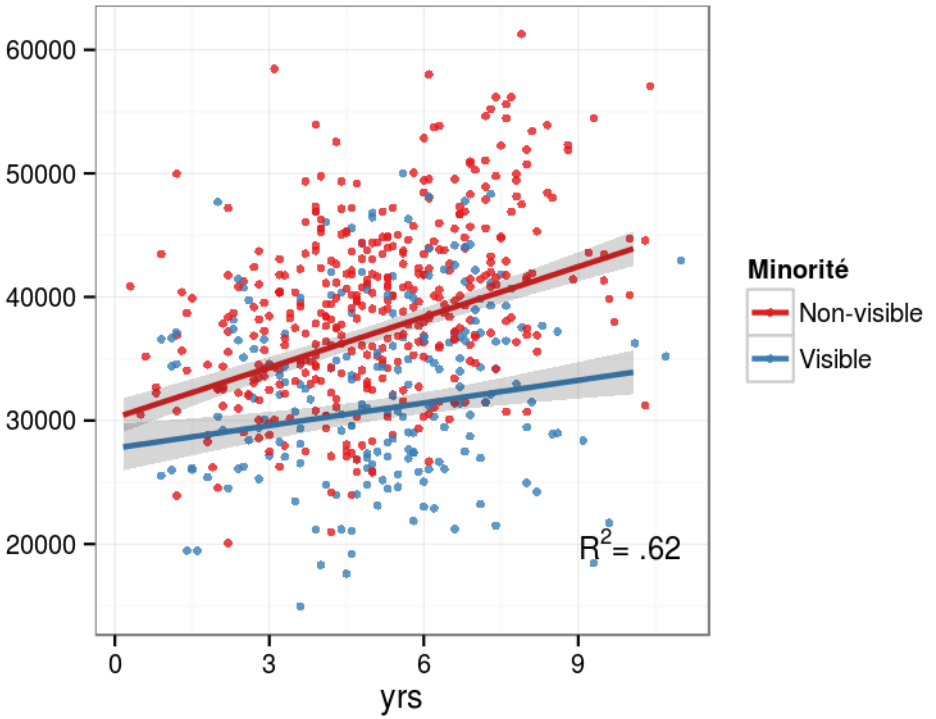
- Globalement: theme_set()
- Localement: p + theme()

theme (localement)

+ theme_grey()

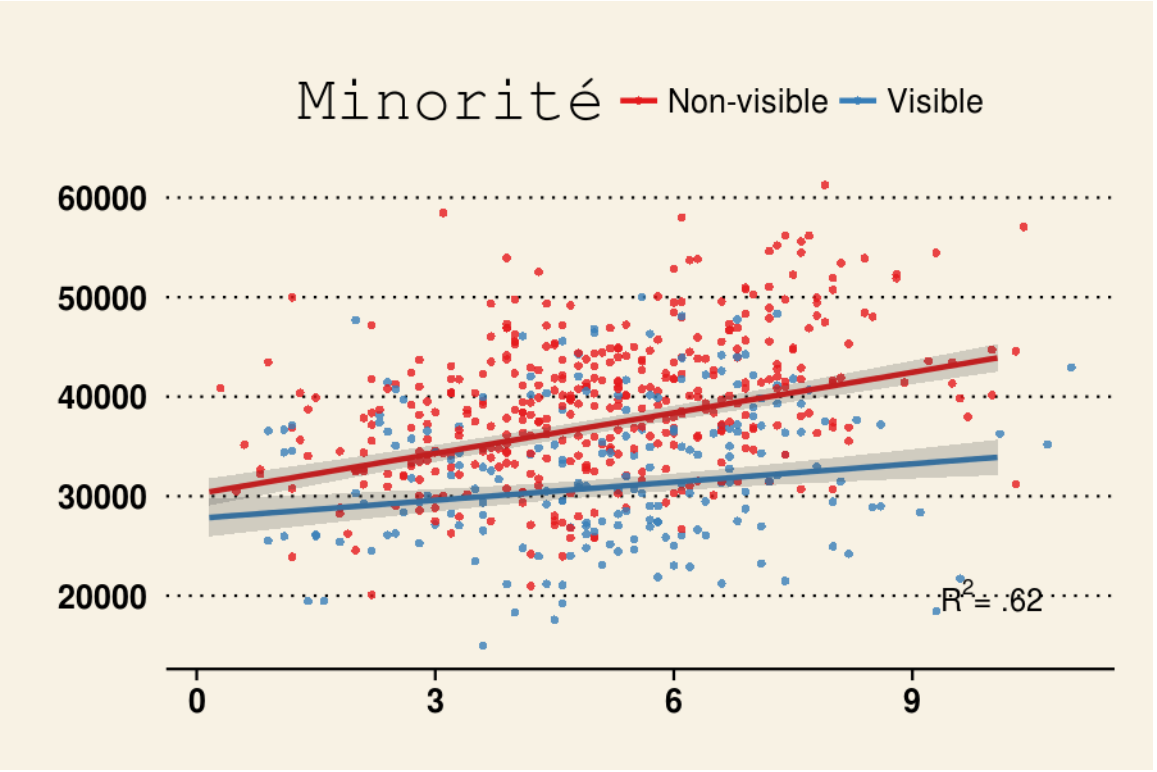


+ theme_bw()



theme (localement)

```
> library(ggthemes)
> t + theme_wsj()
```

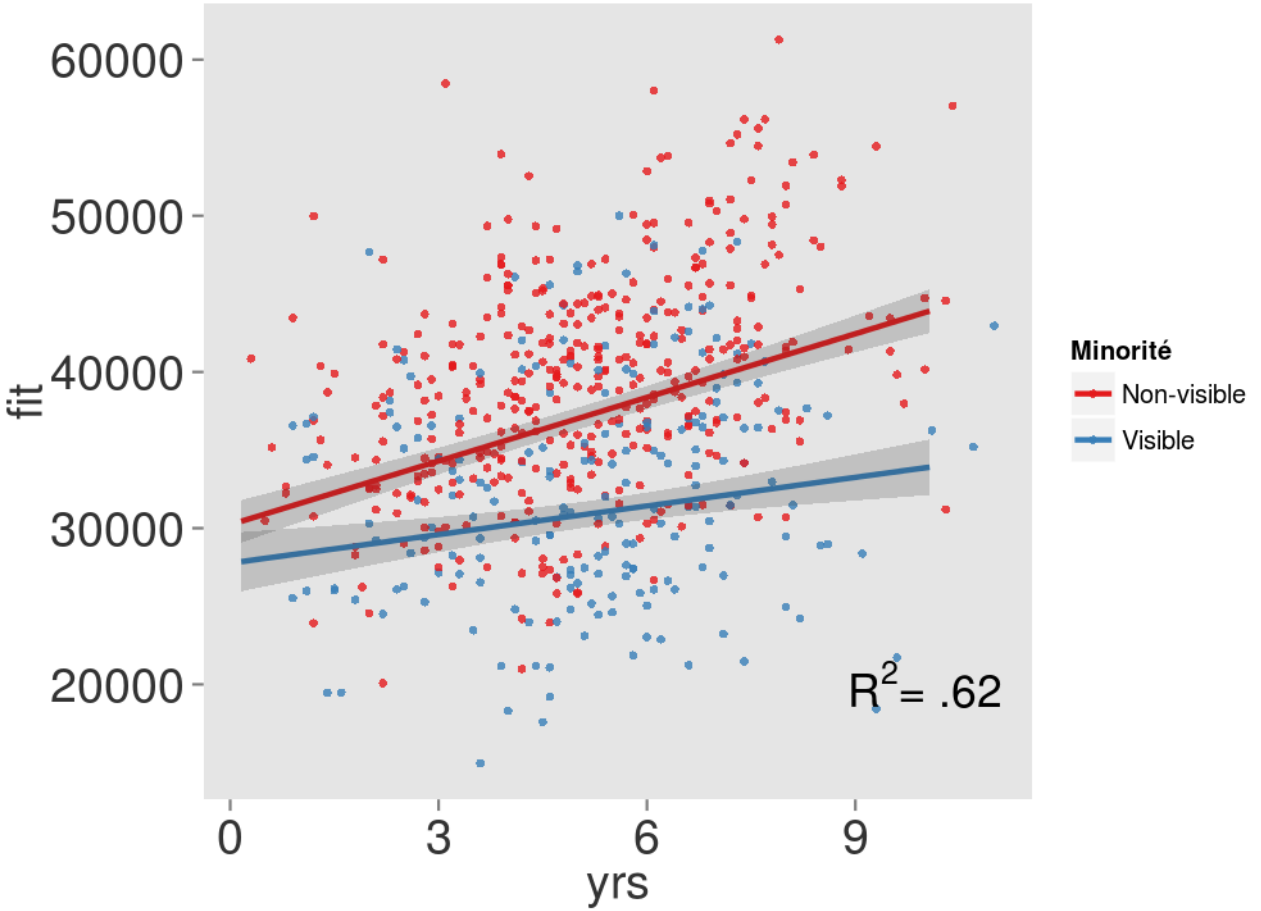


theme (globalement)

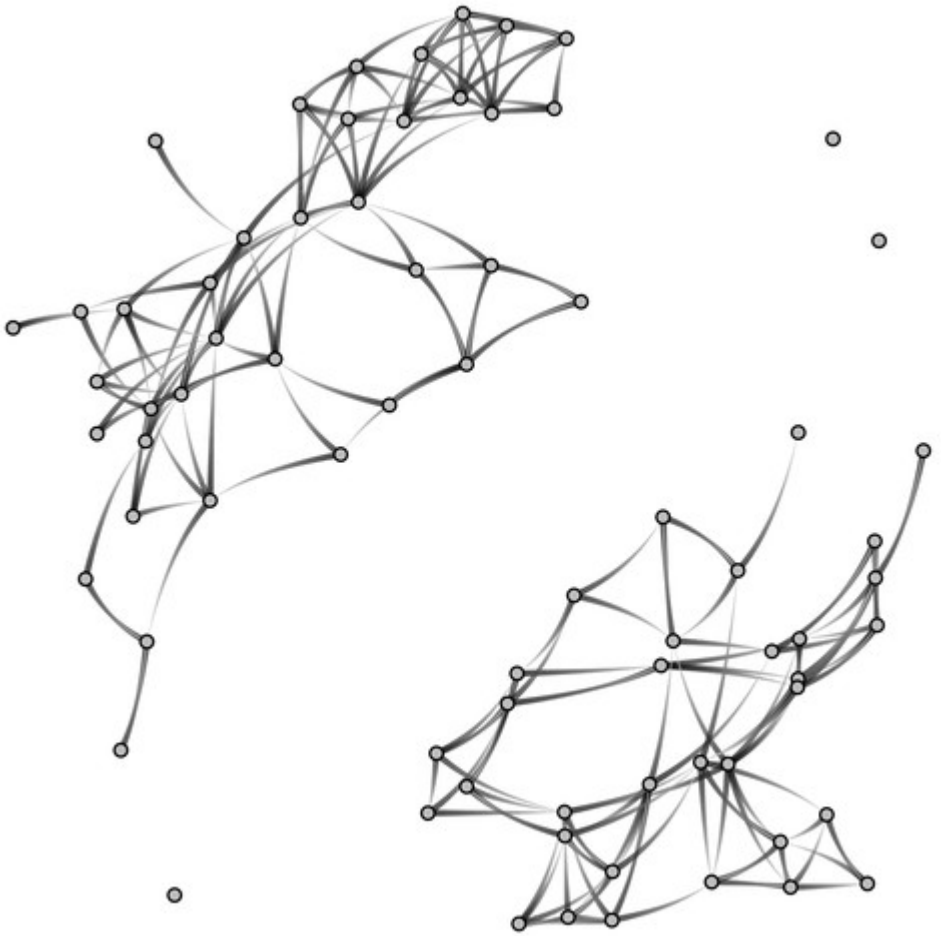
```
> montheme <- theme_gray() +
              theme(panel.grid.major=element_blank(),
                    panel.grid.minor=element_blank(),
                    axis.text=element_text(colour="grey20",
                                           size=rel(1.5)),
                    axis.title=element_text(colour="grey20",
                                           size=rel(1.5)))
```

theme (globalement)

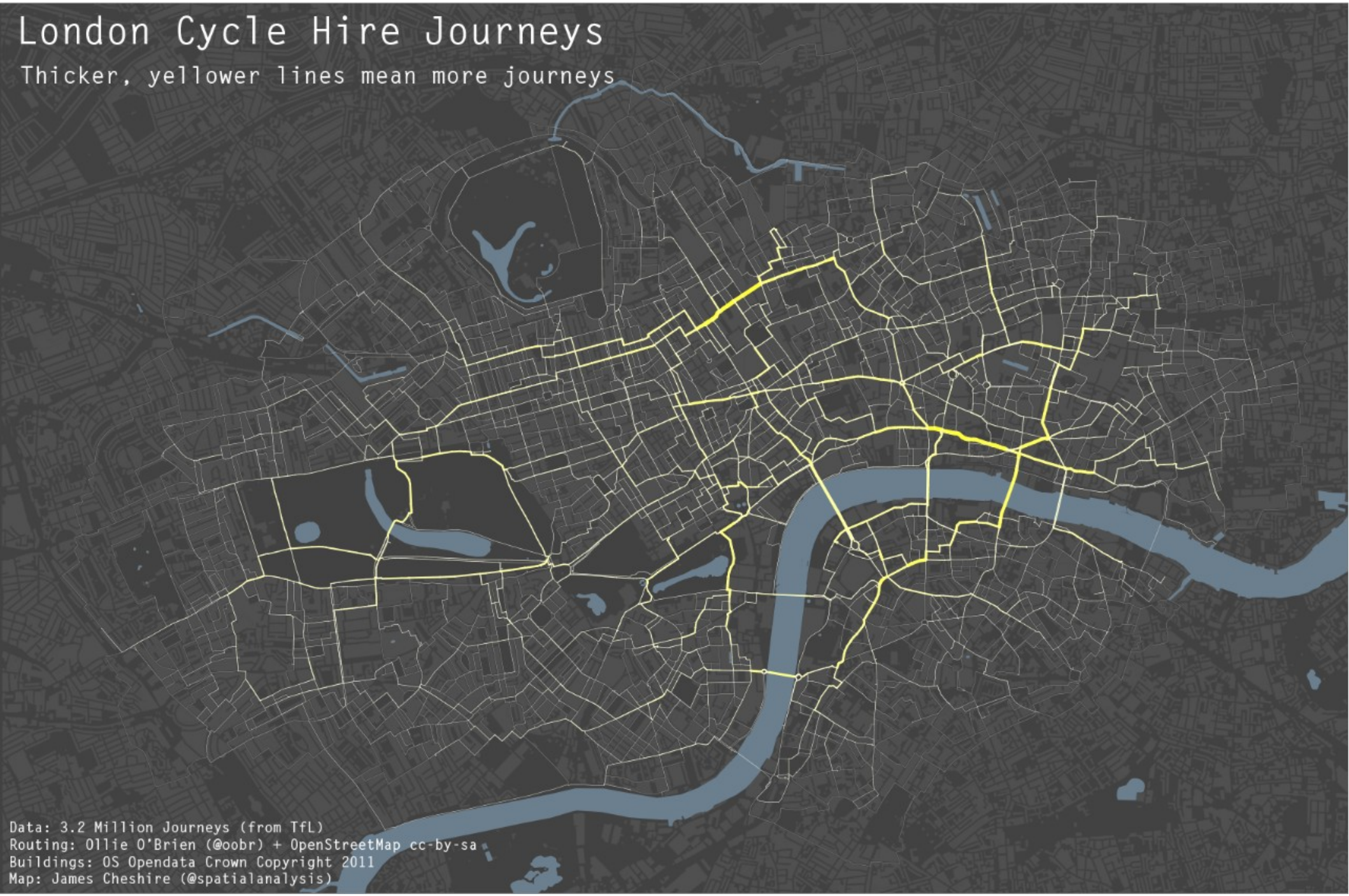
```
> theme_set(montheme)
> t
```



Quelques exemples



http://is-r.tumblr.com/post/38459242505/beautiful-network-diagrams-with-ggplot2


London Cycle Hire Journeys
Thicker, yellower lines mean more journeys
Data: 3.2 Million Journeys (from TfL)
Routing: Ollie O'Brien (@oobr) + OpenStreetMap cc-by-sa
Buildings: OS Opendata Crown Copyright 2011
Map: James Cheshire (@spatialanalysis)

Merci.