

Visualizing data from epidemiologic studies: An expanded scatter plot matrix

Benjamin Barnes, Karen Steindorf

German Cancer Research Center (DKFZ)

Unit of Environmental Epidemiology



2008 Dortmund, 13 August 2008

b.barnes@dkfz.de

dkfz.

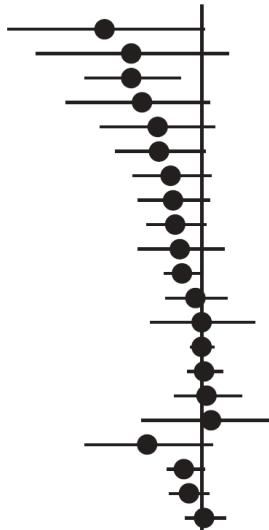
GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION

Physical Activity and Breast Cancer

Relative risk, high vs. low physical activity

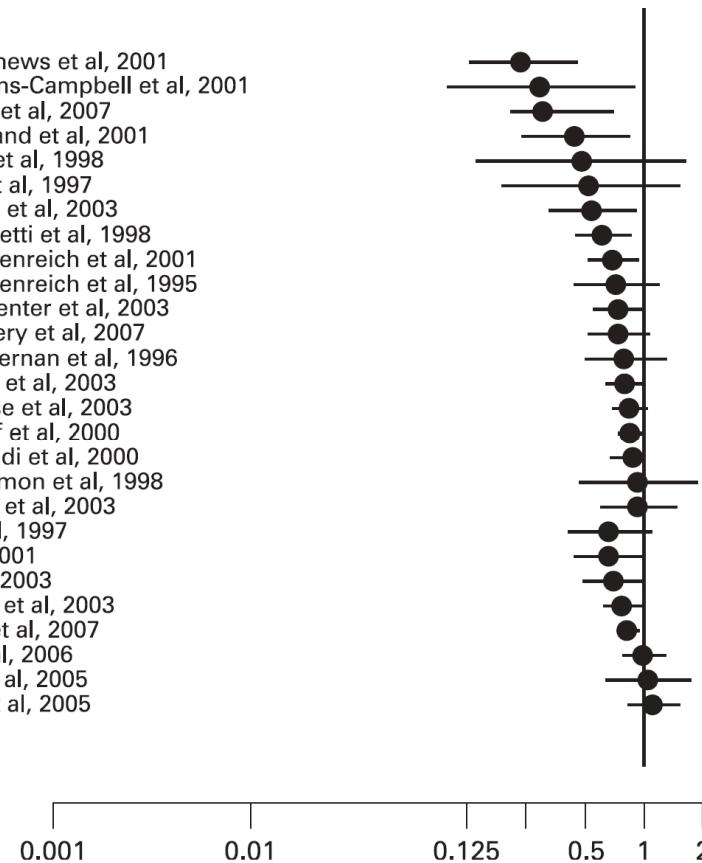
Premenopausal

- Case-control: Ueji et al, 1998
- Case-control: Kruk et al, 2007
- Case-control: Yang et al, 2003
- Case-control: Adams-Campbell et al, 2001
- Case-control: Friedenreich et al, 1995
- Case-control: Matthews et al, 2001
- Case-control: Slattery et al, 2007
- Case-control: Mezzetti et al, 1998
- Case-control: John et al, 2003
- Case-control: Dorn et al, 2003
- Case-control: Hirose et al, 2003
- Case-control: Steindorf et al, 2003
- Case-control: Hu et al, 1997
- Case-control: Magnusson et al, 2005
- Case-control: Gammon et al, 1998
- Case-control: Friedenreich et al, 2001
- Case-control: Gilliland et al, 2001
- Cohort: Thune et al, 1997
- Cohort: Lahmann et al, 2007
- Cohort: Silvera et al, 2006
- Cohort: Colditz et al, 2003



Postmenopausal

- Case-control: Matthews et al, 2001
- Case-control: Adams-Campbell et al, 2001
- Case-control: Kruk et al, 2007
- Case-control: Gilliland et al, 2001
- Case-control: Ueji et al, 1998
- Case-control: Hu et al, 1997
- Case-control: Yang et al, 2003
- Case-control: Mezzetti et al, 1998
- Case-control: Friedenreich et al, 2001
- Case-control: Friedenreich et al, 1995
- Case-control: Carpenter et al, 2003
- Case-control: Slattery et al, 2007
- Case-control: McTiernan et al, 1996
- Case-control: John et al, 2003
- Case-control: Hirose et al, 2003
- Case-control: Shoff et al, 2000
- Case-control: Moradi et al, 2000
- Case-control: Gammon et al, 1998
- Case-control: Dorn et al, 2003
- Cohort: Thune et al, 1997
- Cohort: Lee et al, 2001
- Cohort: Patel et al, 2003
- Cohort: McTiernan et al, 2003
- Cohort: Lahmann et al, 2007
- Cohort: Silvera et al, 2006
- Cohort: Mertens et al, 2005
- Cohort: Schnohr et al, 2005



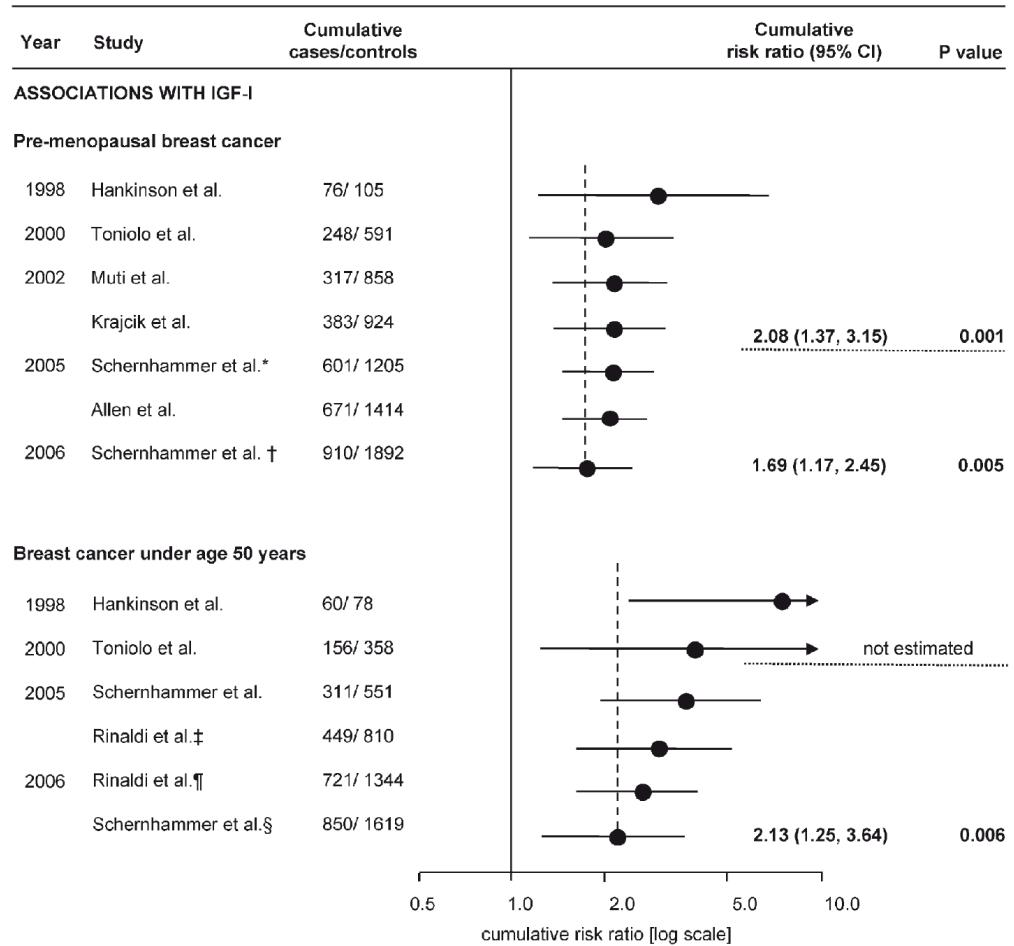
Friedenreich C, Cust A. Physical activity and breast cancer risk: impact of timing, type and dose of activity and population subgroup effects. *Br J Sport Med* (2008). [Epub ahead of print].

Hypotheses how physical activity might affect cancer risk

- Sex steroid hormones
- Insulin and glucose
- Immune system
- Inflammatory factors
- Insulin-like growth factor (IGF) system

IGF-I and Breast Cancer

- IGF-I is mitogenic and antiapoptotic
- High IGF-I levels associated with cancer risk (breast, colon, prostate)



Renehan A, Harvie M, Howell A. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and breast cancer risk: eight years on. *Endocrine-Related Cancer* (2006); 13: 273-278.

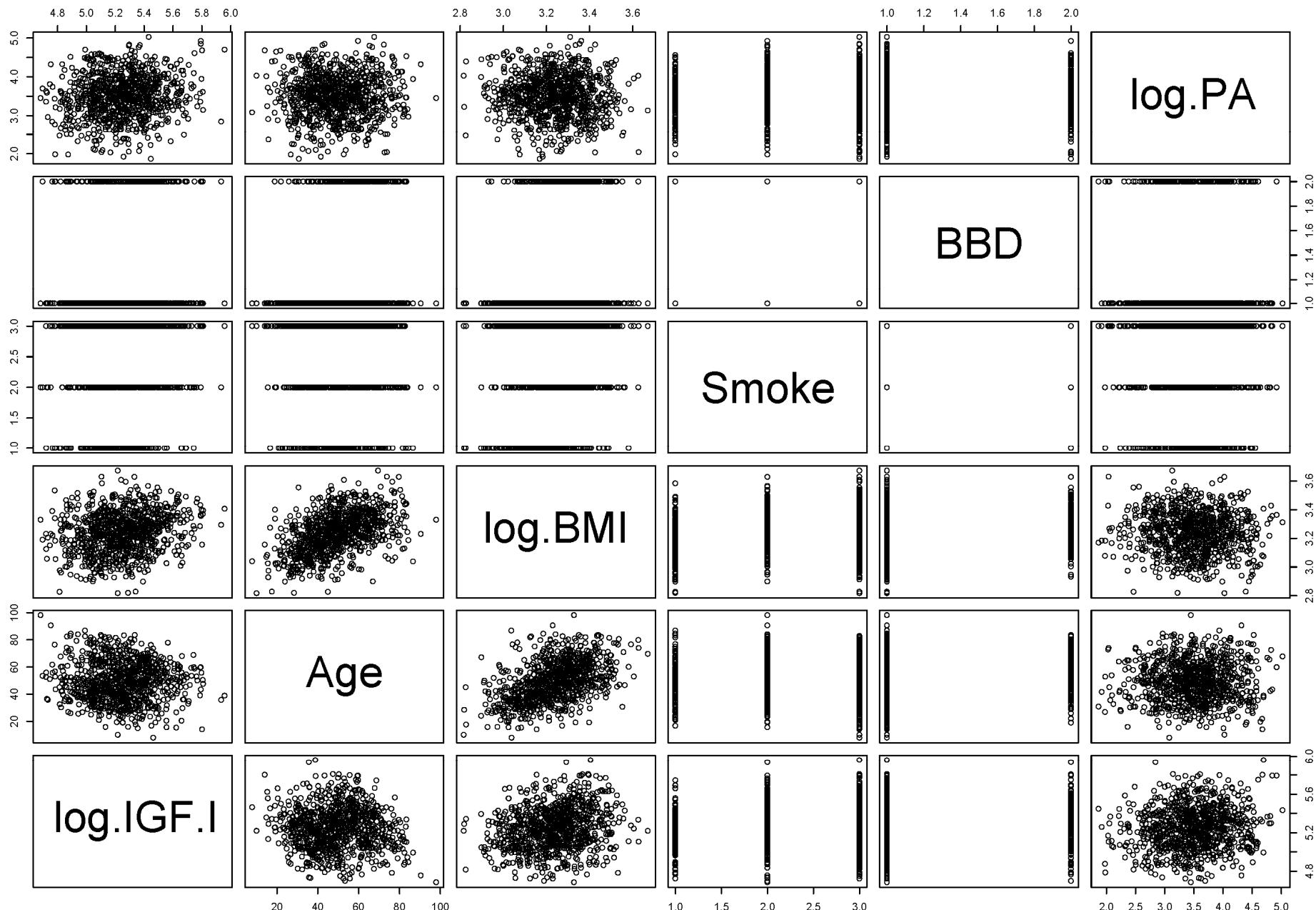
What might affect IGF-I levels?

- Non-modifiable
 - Age
 - Benign Breast disease
- Modifiable
 - Body Mass Index
 - Smoking
 - Physical activity

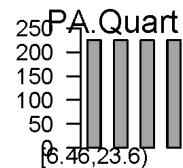
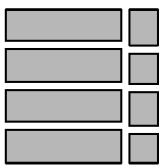
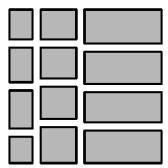
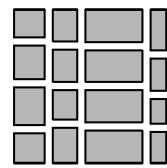
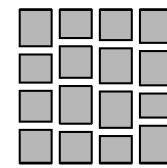
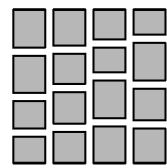
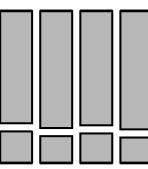
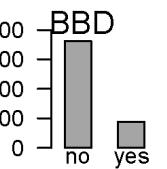
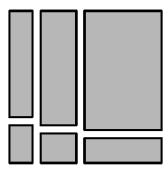
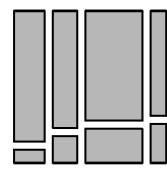
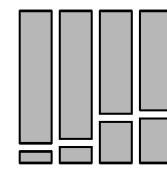
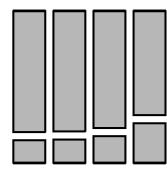
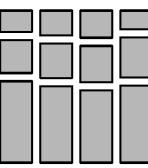
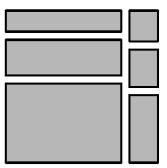
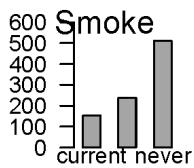
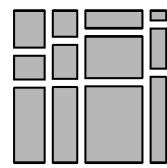
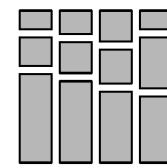
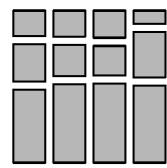
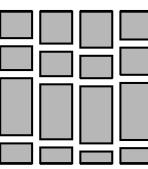
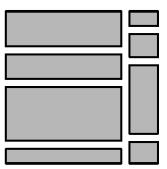
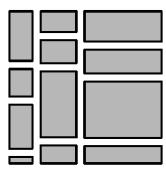
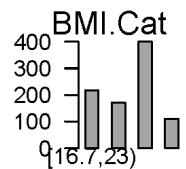
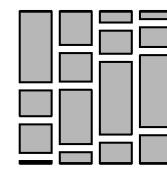
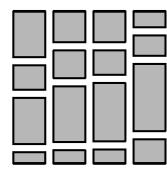
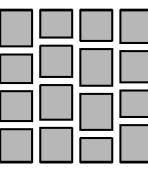
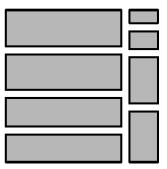
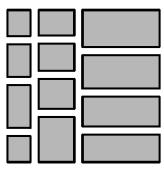
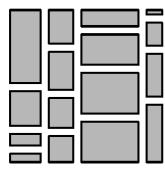
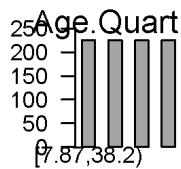
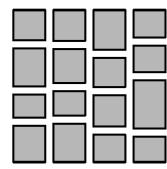
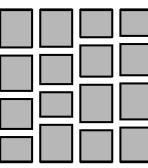
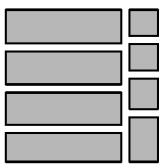
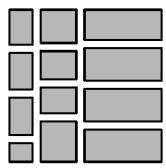
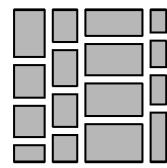
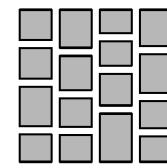
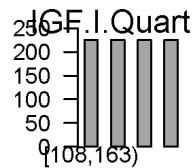
Continuous | Categorical

How can this data be visualized for EDA?

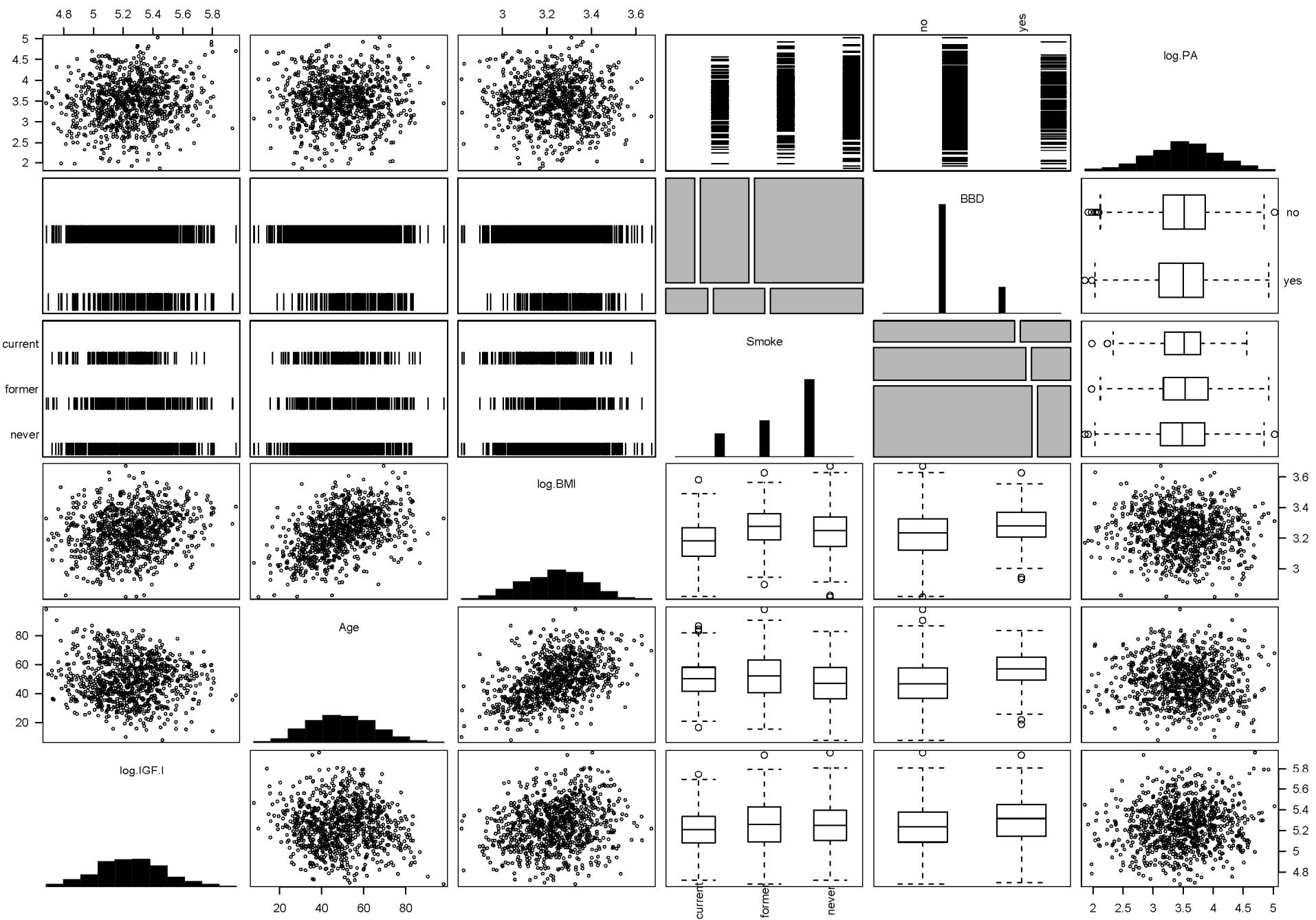
pairs()



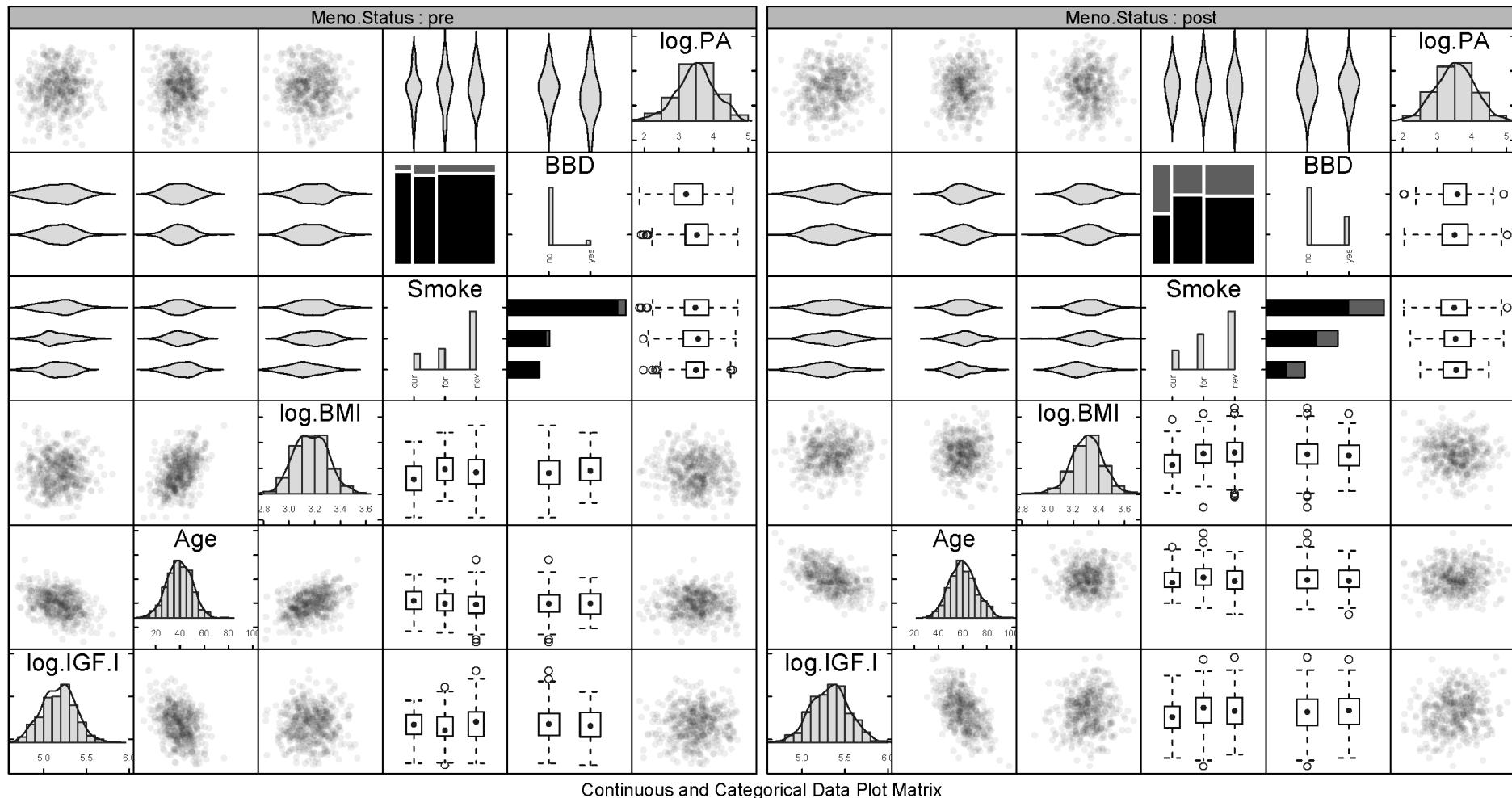
pairs.table()



gpairs()



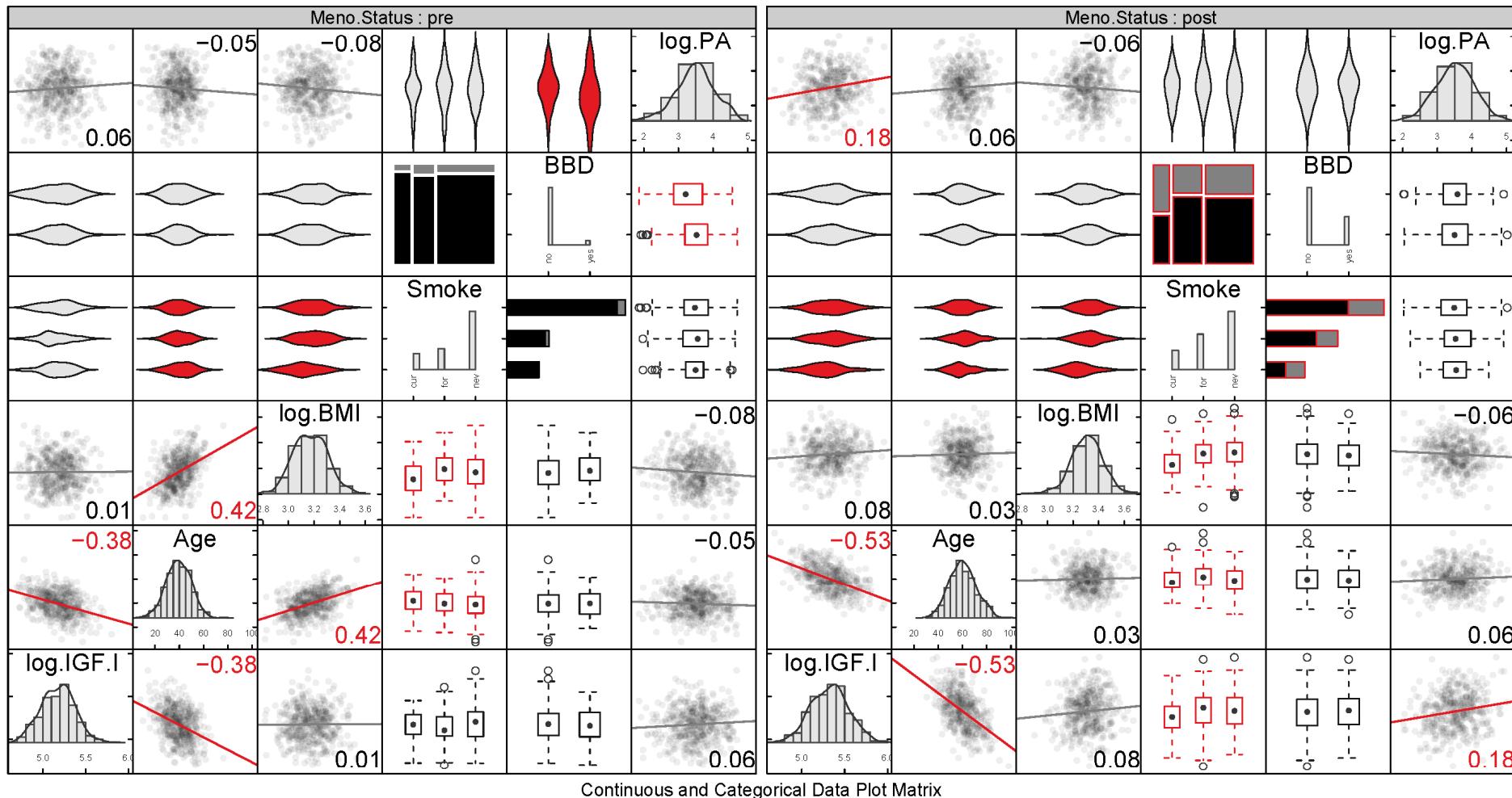
```
my.splom(input.vars, cond.var = "Meno.Status", data, sig.col = "red",
  stats = FALSE, xlab = "Continuous and Categorical Data Plot Matrix",
  alpha = 0.05, , upper.plots = NULL, lower.plots = NULL ...)
```



```

my.splom(input.vars, cond.var = "Meno.Status", data, sig.col = "red",
          stats = "bivar", xlab = "Continuous and Categorical Data Plot Matrix",
          alpha = 0.05, , upper.plots = NULL, lower.plots = NULL ...)

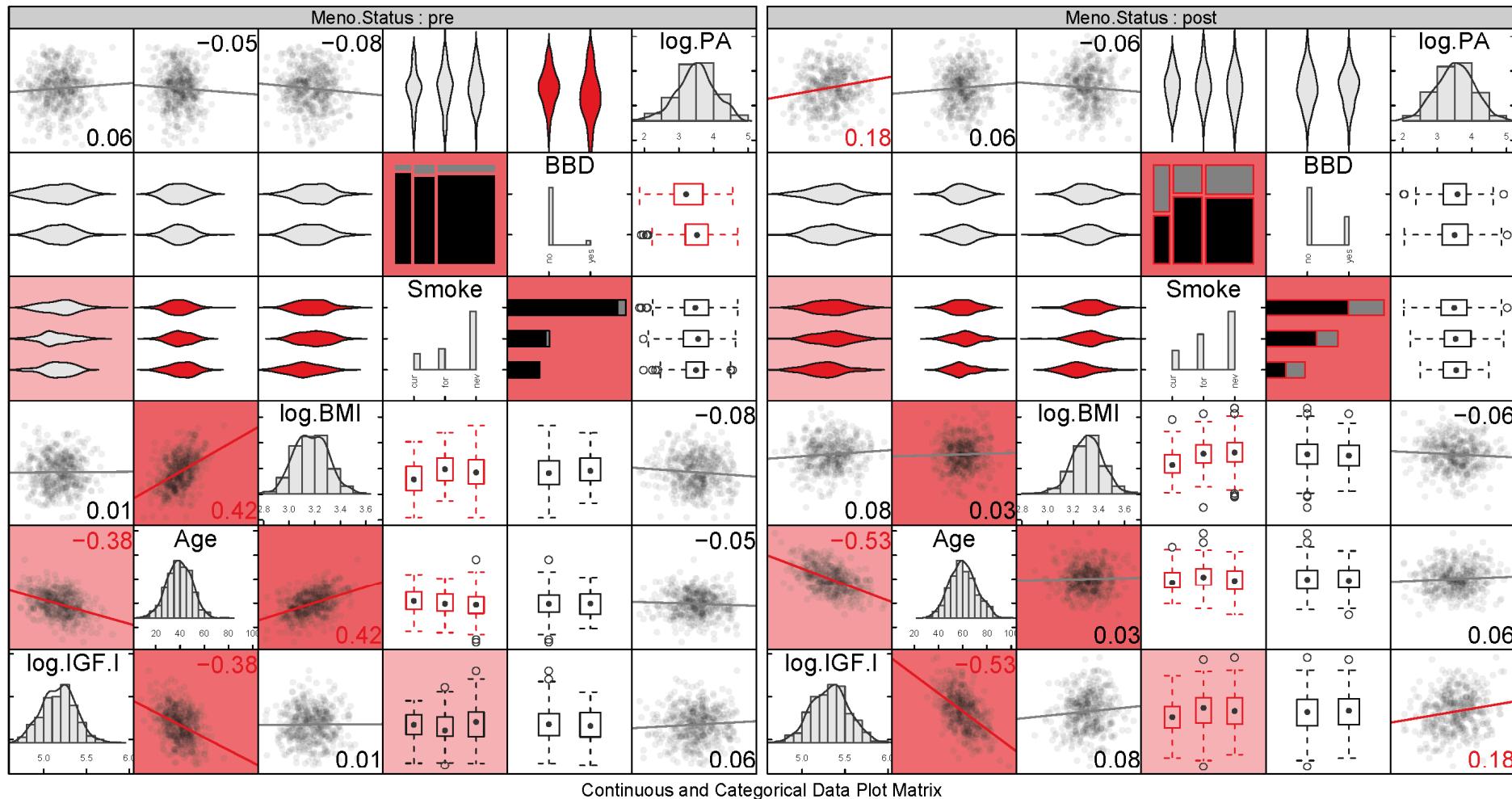
```



```

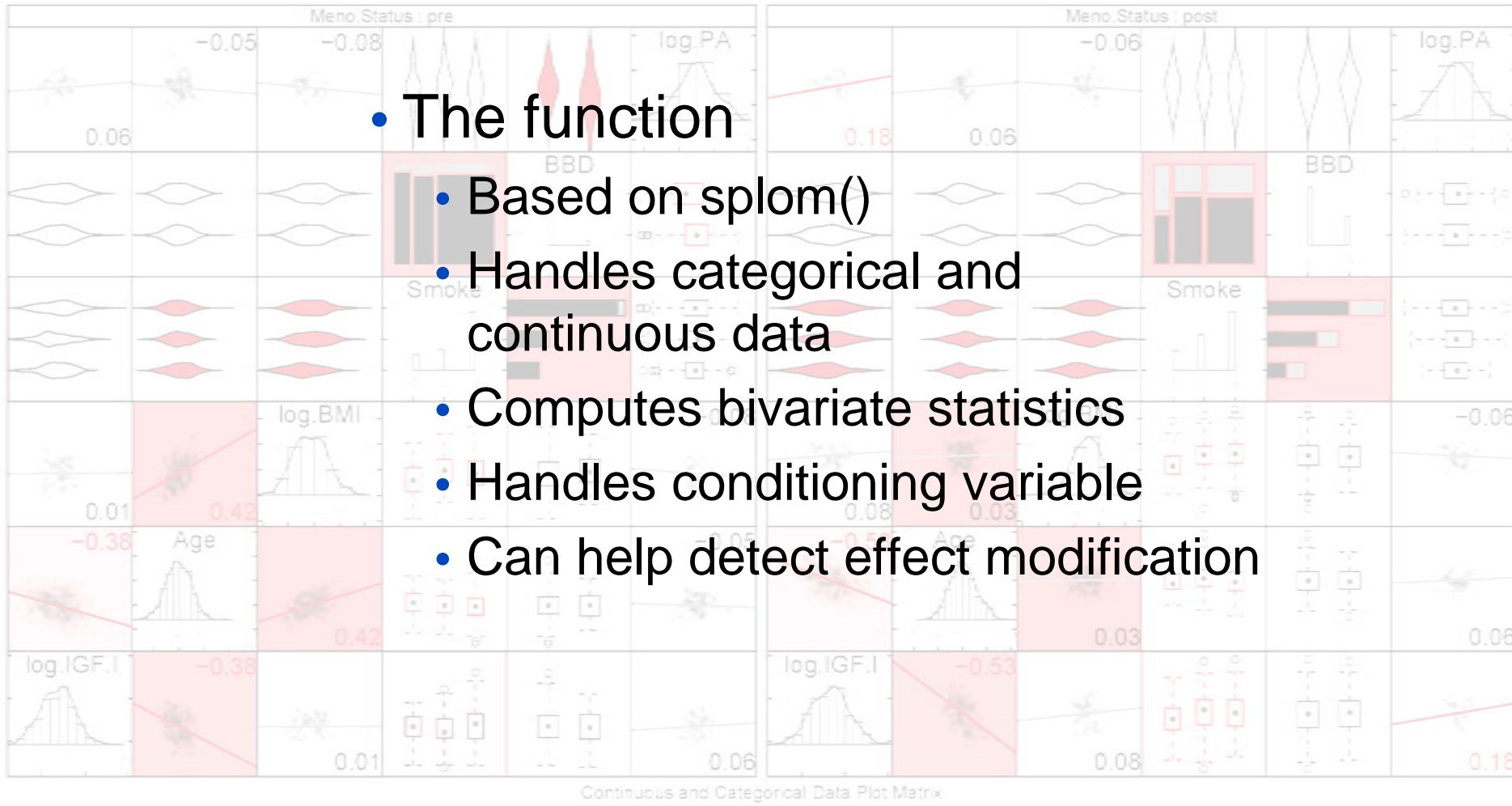
my.splom(input.vars, cond.var = "Meno.Status", data, sig.col = "red",
  stats = c("bivar", "int"), xlab = "Continuous and Categorical Data Plot
Matrix", alpha = 0.05, , upper.plots = NULL, lower.plots = NULL ...)

```



Conclusions I

- The function
 - Based on `splom()`
 - Handles categorical and continuous data
 - Computes bivariate statistics
 - Handles conditioning variable
 - Can help detect effect modification



Conclusions II

