Visualizing data from epidemiologic studies:
An expanded scatter plot matrix

Benjamin Barnes, Karen Steindorf

German Cancer Research Center (DKFZ)

Unit of Environmental Epidemiology

useR! 2008 Dortmund, 13 August 2008

b.barnes@dkfz.de
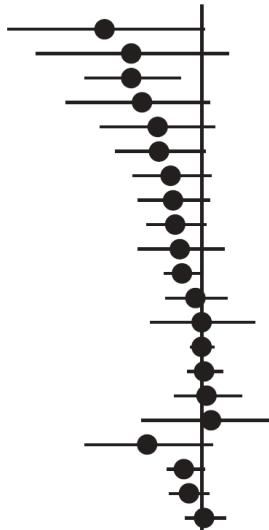
dkfz.
GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION

Physical Activity and Breast Cancer

## Relative risk, high vs. low physical activity

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.
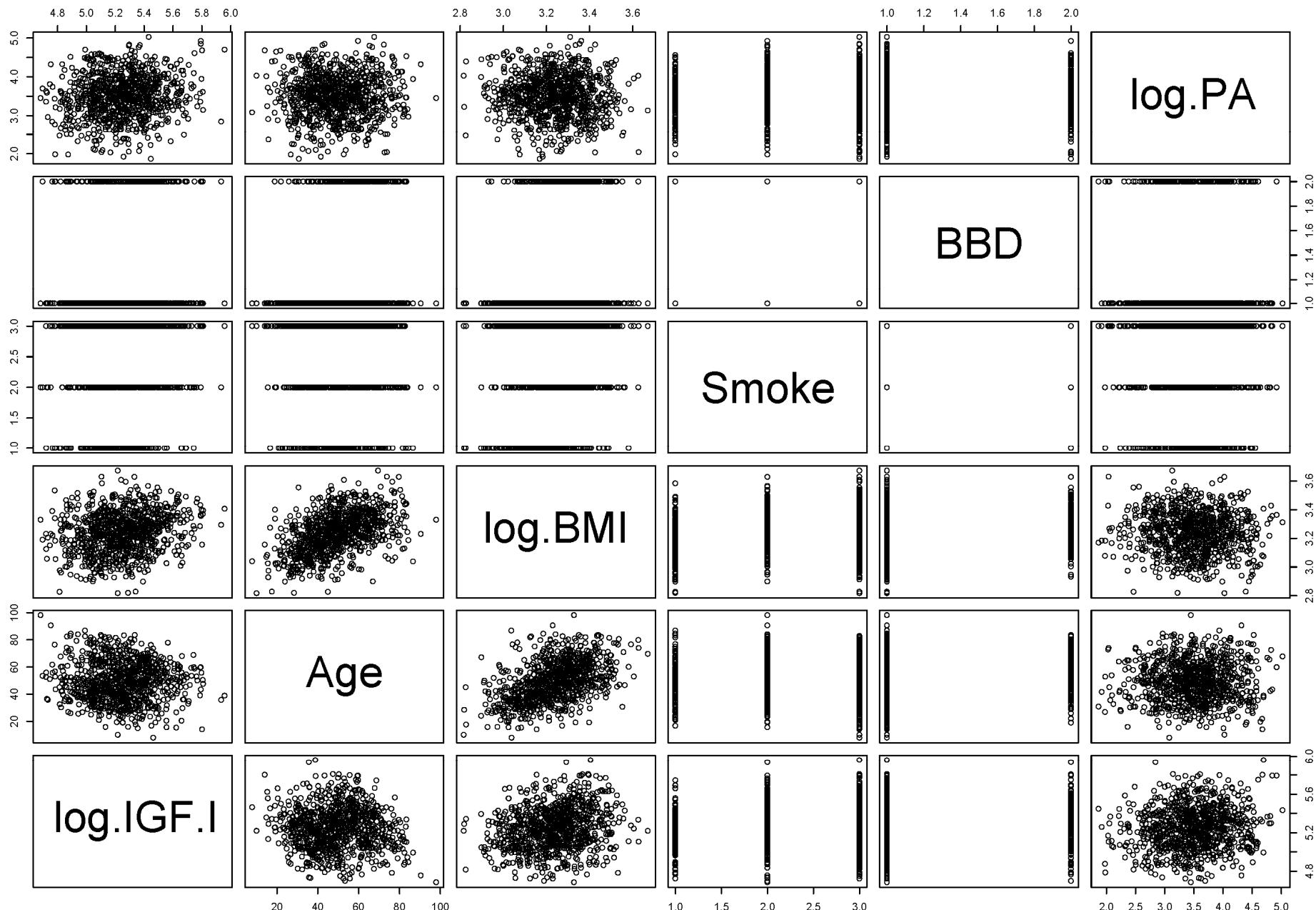
What might affect IGF-I levels?

- Non-modifiable
  - **Age**
  - **Benign Breast disease**

- Modifiable
  - **Body Mass Index**
  - **Smoking**
 - **Physical activity**
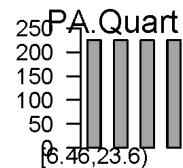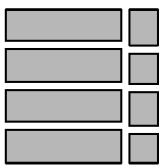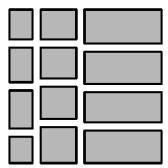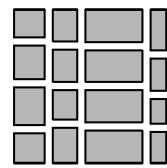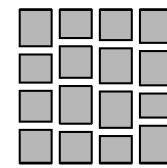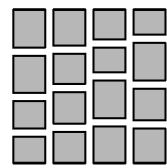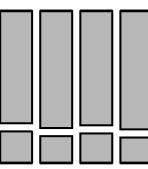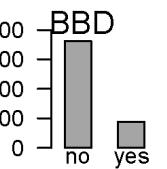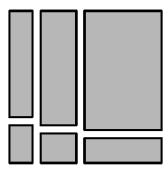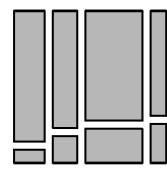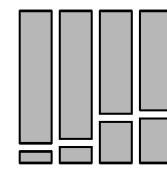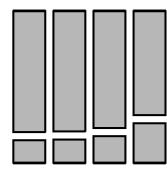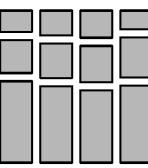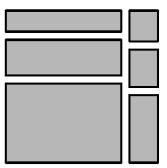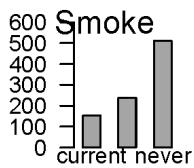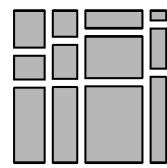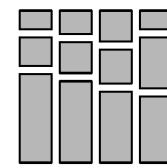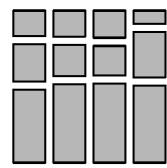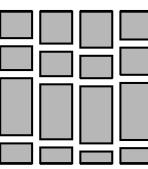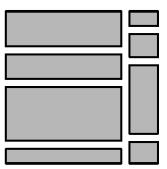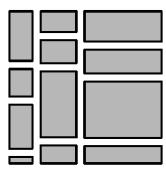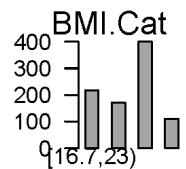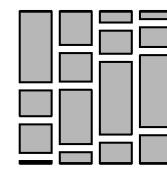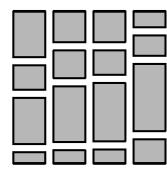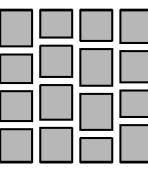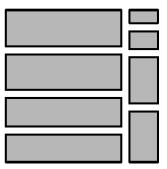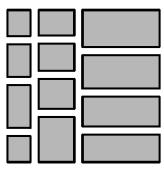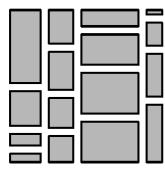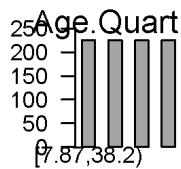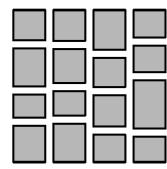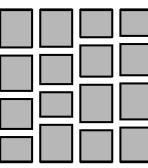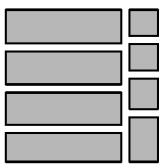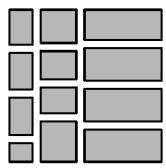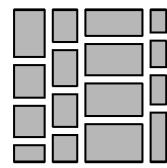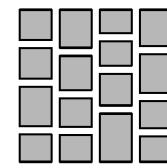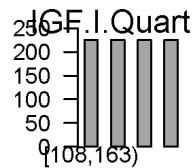
Continuous | Categorical

How can this data be visualized for EDA?
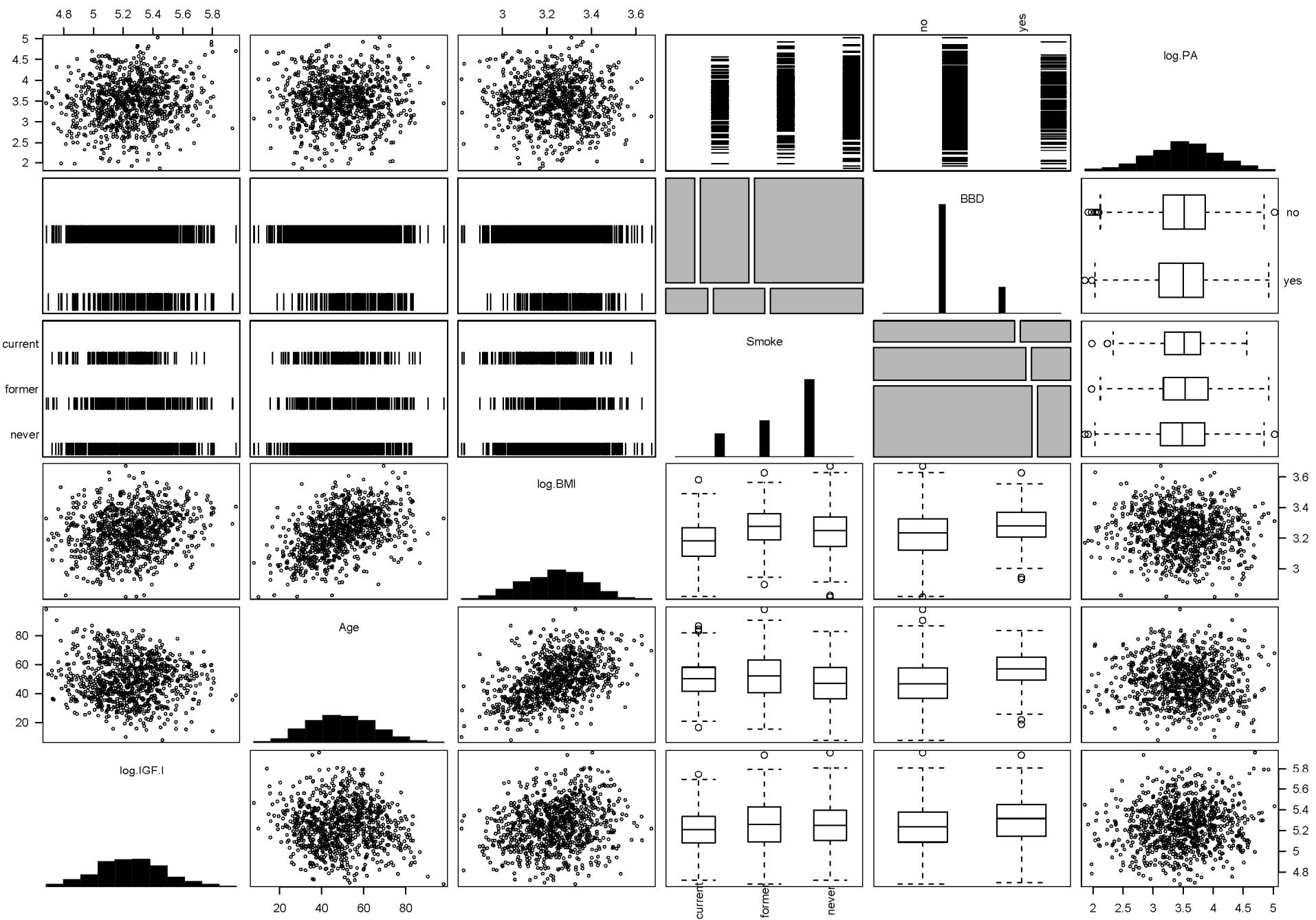
pairs()
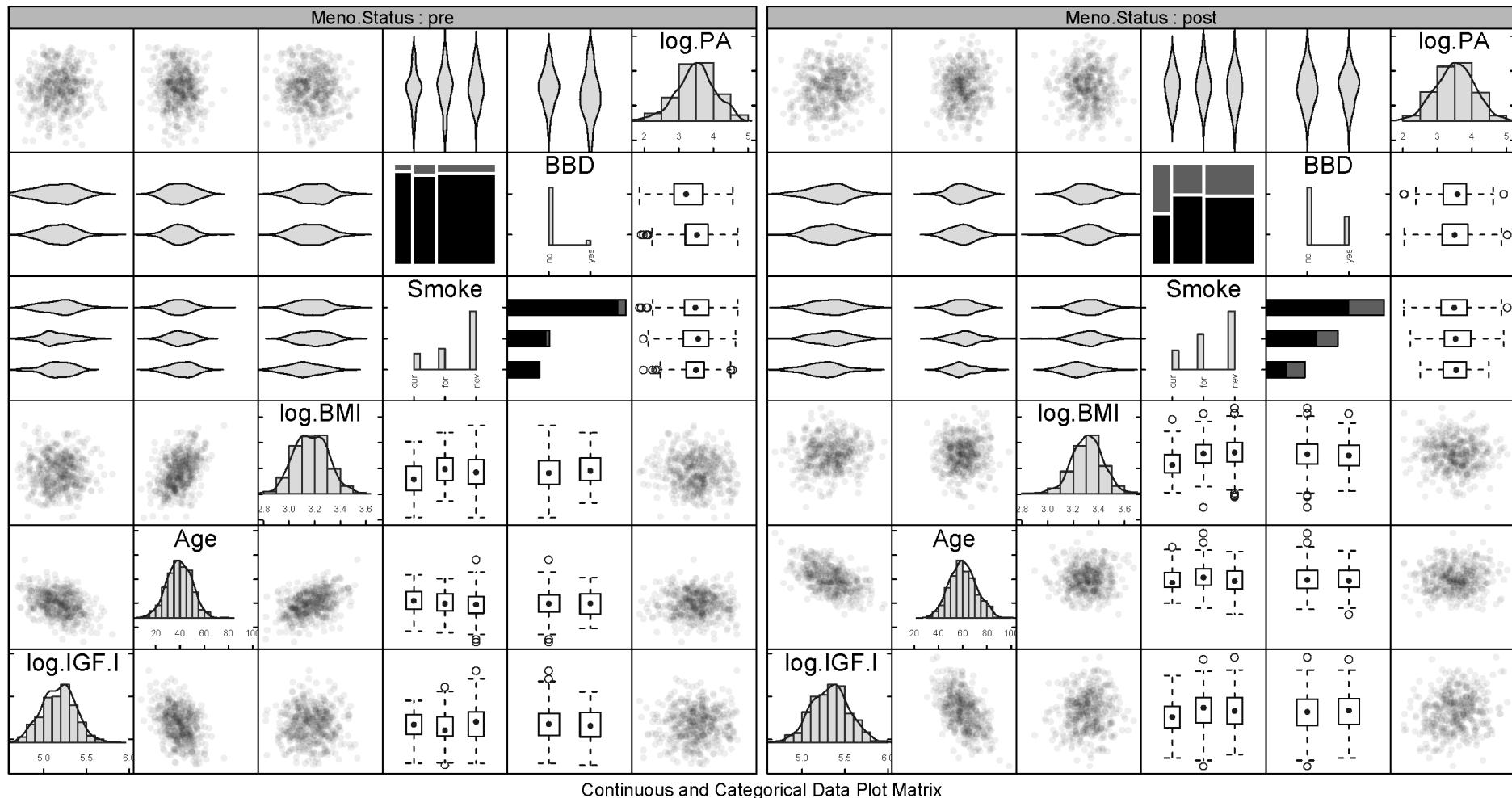
pairs.table()

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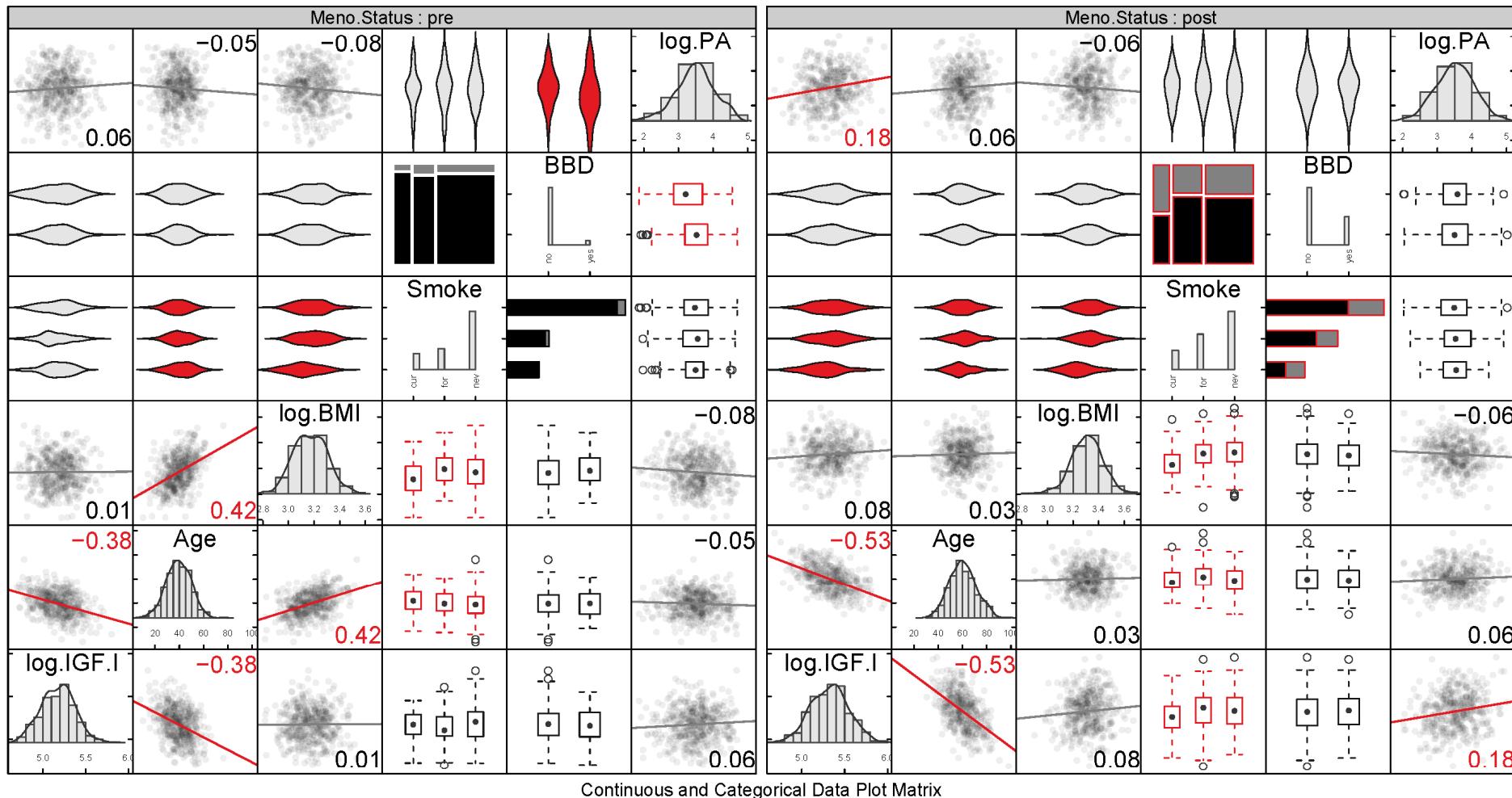
my.splom(input.vars, cond.var = "Meno.Status", data, sig.col = "red",
    stats = FALSE, xlab = "Continuous and Categorical Data Plot Matrix",
   alpha = 0.05, , upper.plots = NULL, lower.plots = NULL ...)

Continuous and Categorical Data Plot Matrix

```
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 ...)
```


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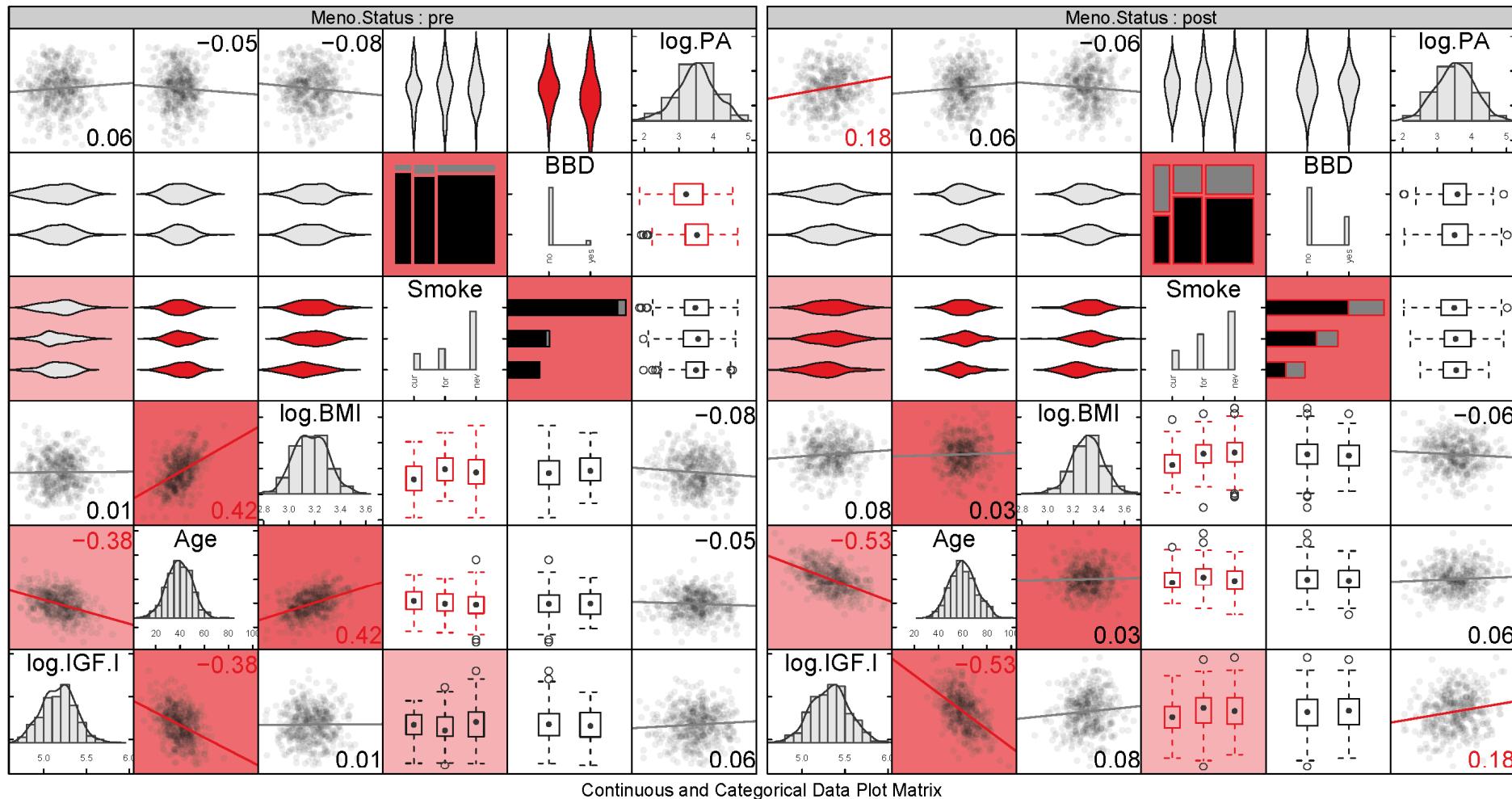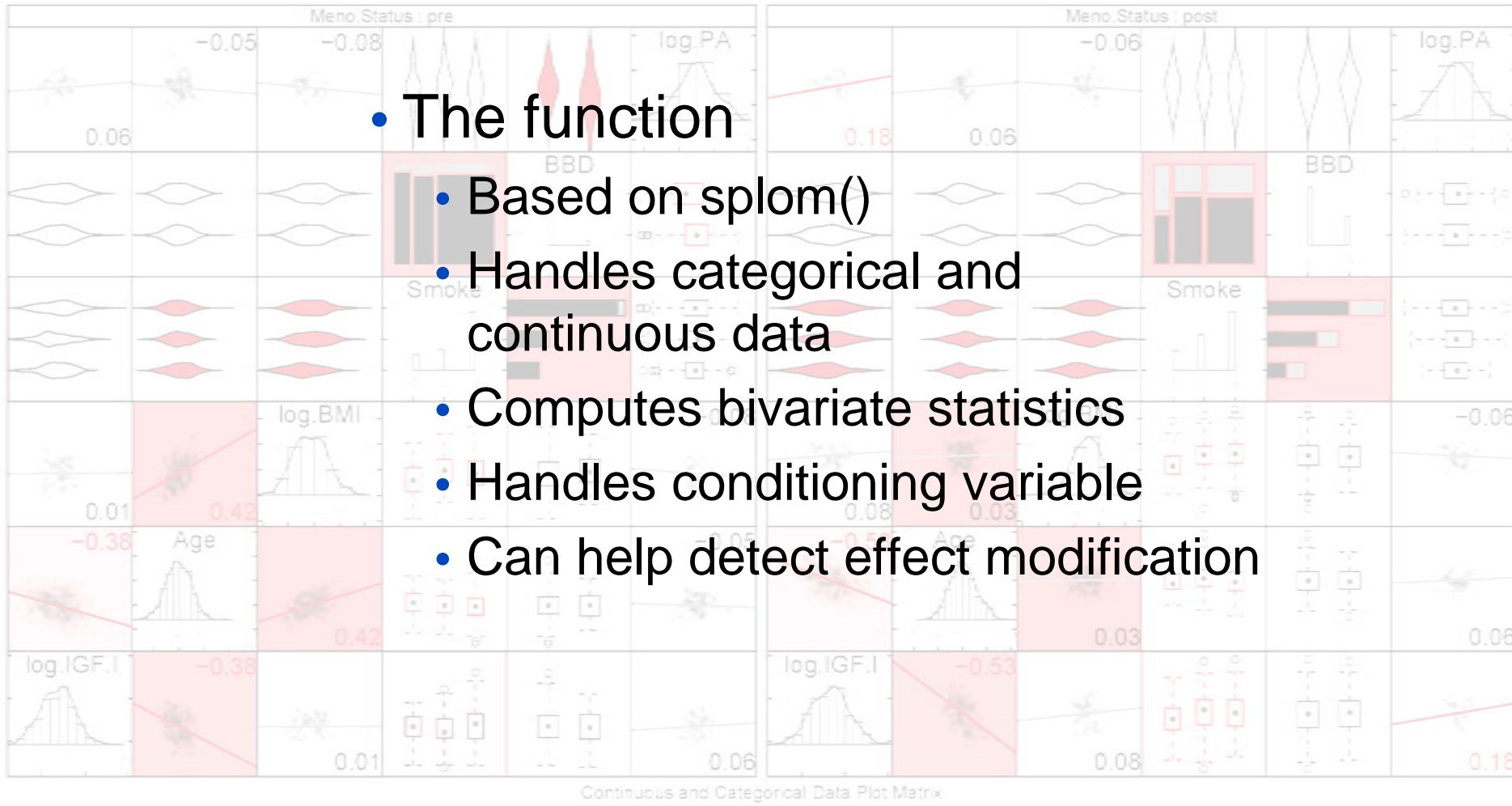
my.splom(input.vars, cond.var = "Meno.Status", data, sig.col = "red",
stats = **c("bivar","int")**, xlab = "Continuous and Categorical Data Plot
Matrix", alpha = 0.05, , upper.plots = NULL, lower.plots = NULL ...)


Continuous and Categorical Data Plot Matrix

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

Continuous and Categorical Data Plot Matrix