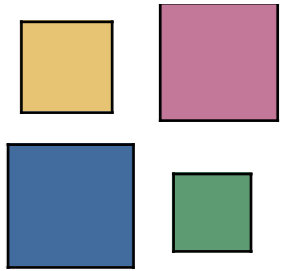


## Plot Ensembles

- **Idee**

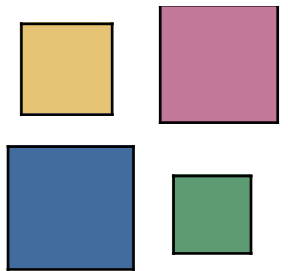
Bestimmte Fragestellungen treten in der EDA immer wieder auf, und werden meist mit den gleichen Tools gelöst  $\Rightarrow$  Formalisierung

- Strategien der ISG können im ersten Schritt am einfachsten von Methoden der klassischen Statistik abgeleitet werden.
- Einige Techniken sind direkt im Plot verfügbar, z.B. Spineplot etc.
- Andere findet man in speziellen Plot Ensembles, welche bestimmte Muster visualisieren können.
- Um eine größere statistische Sicherheit zu bekommen, kann man statistische Graphik mit Schätzern und Tests anreichern.
- Strategien sollten aber niemals zu schematisch werden:  
 $\Rightarrow$  Die Exploration ist das mächtigste Werkzeug der stat. Graphik!



## Enhanced Graphics

- Die Vorteile, statistische Graphiken mit Schätzern und Tests anzureichern liegen auf der Hand.
- Die jeweiligen Voraussetzungen der Verfahren sind jedoch nur sehr schwer zu überprüfen  $\Rightarrow$  nichtparametrische Verfahren sind oftmals die bessere, robustere Wahl.
- Das Problem des “multiplen Testens” ist in einem interaktiven System natürlich immanent.
- Werden Parameter wie Bandbreiten eingebracht, so ist meist nicht klar, wie diese auf die selektierte Untergruppe propagiert werden sollen.



## 2 dim. Interaktionen

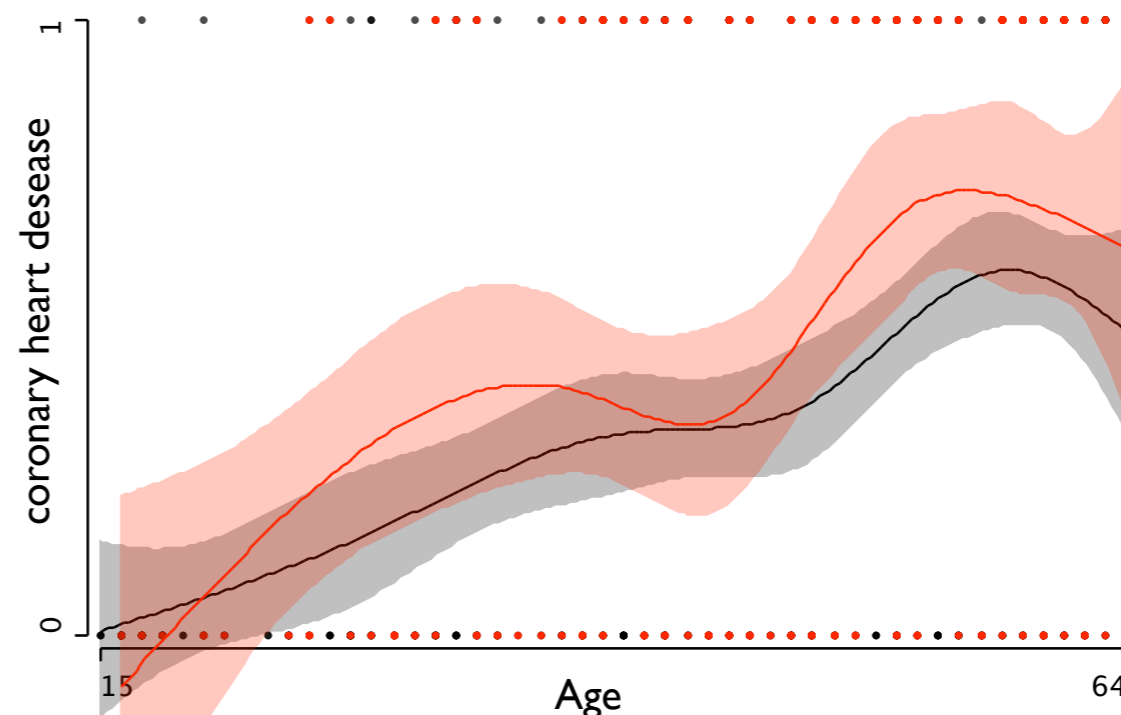
- 2 kategoriale Variablen  $\Rightarrow$  Mosaic Plots ✓

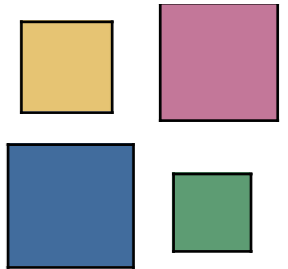
(Assoziationen können über die Darstellung der erwarteten Werte eines Modelles gezeigt werden.)

- 2 stetige Variablen  $\Rightarrow$  Scatterplots ✓

(Falls die Rolle der abhängigen und unabhängigen Variablen bekannt ist, können Scatterplot-Smoother die Graphik ergänzen.)

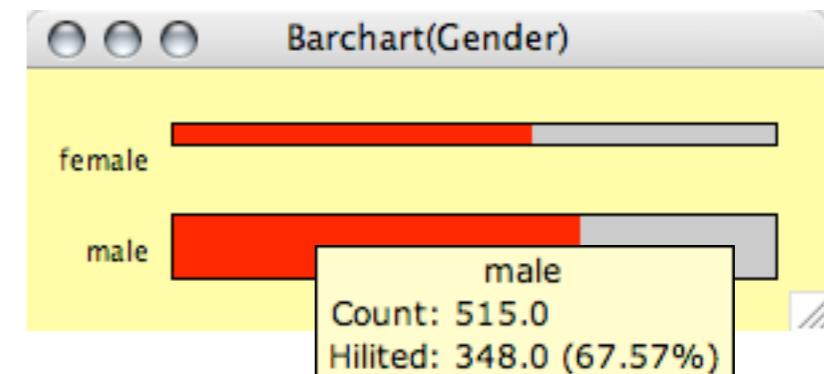
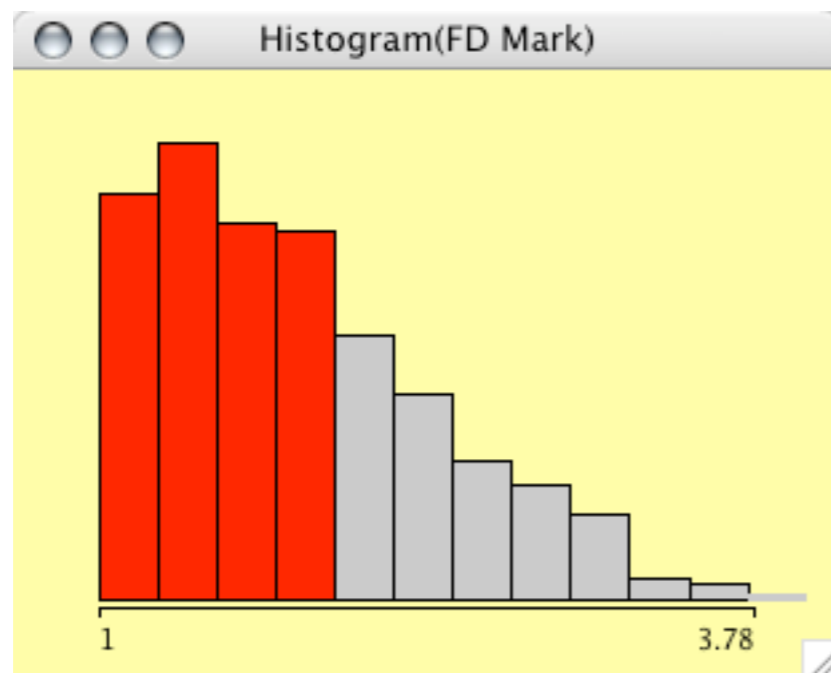
- In Ausnahmefällen kann es auch sinnvoll sein, kategoriale Variablen im Scatterplot darzustellen.





## 2-dim Interaktionen

- Gemischte Skalen
  - stetig  $\Rightarrow$  kategoriell

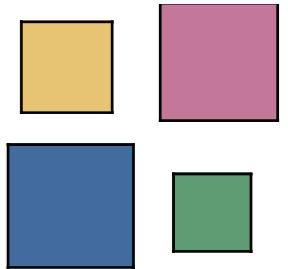


zu lesen:

“gegeben, dass das Geschlecht ‘male’ ist, was ist die Wahrscheinlichkeit, dass die Note besser als 2 ist?”

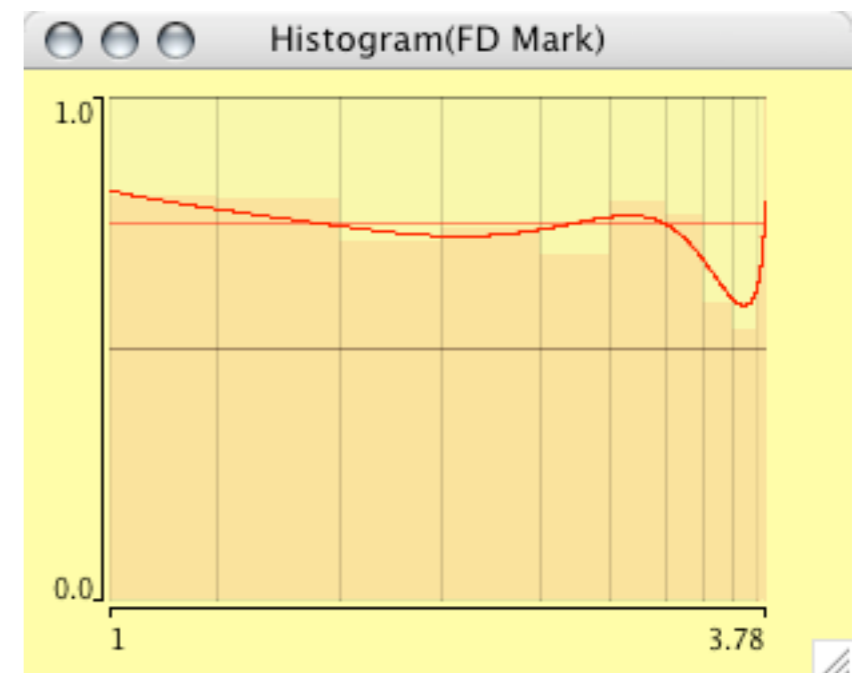
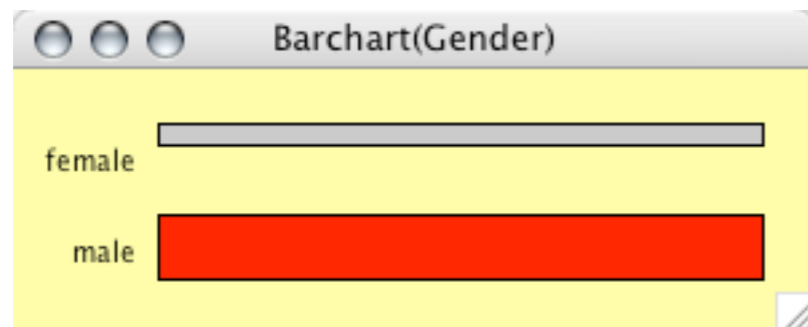
oder

$$P(\text{Note} < 2 \mid \text{Geschlecht} = \text{'male'})$$



## 2-dim Interaktionen

- Gemischte Skalen
  - kategoriell  $\Rightarrow$  stetig



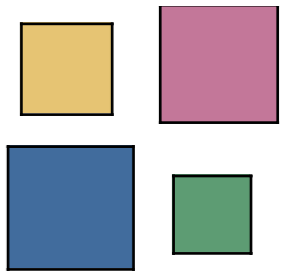
zu lesen:

“für eine Note besser als 1.25, was ist die Wahrscheinlichkeit für ‘male’”

oder

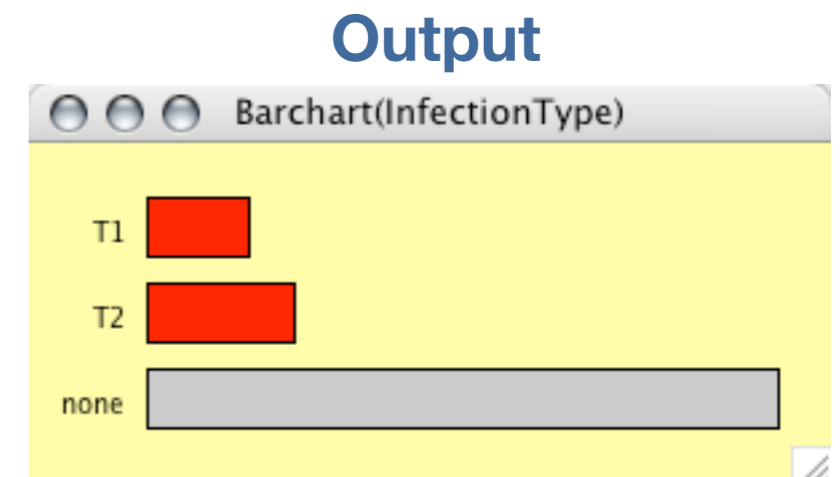
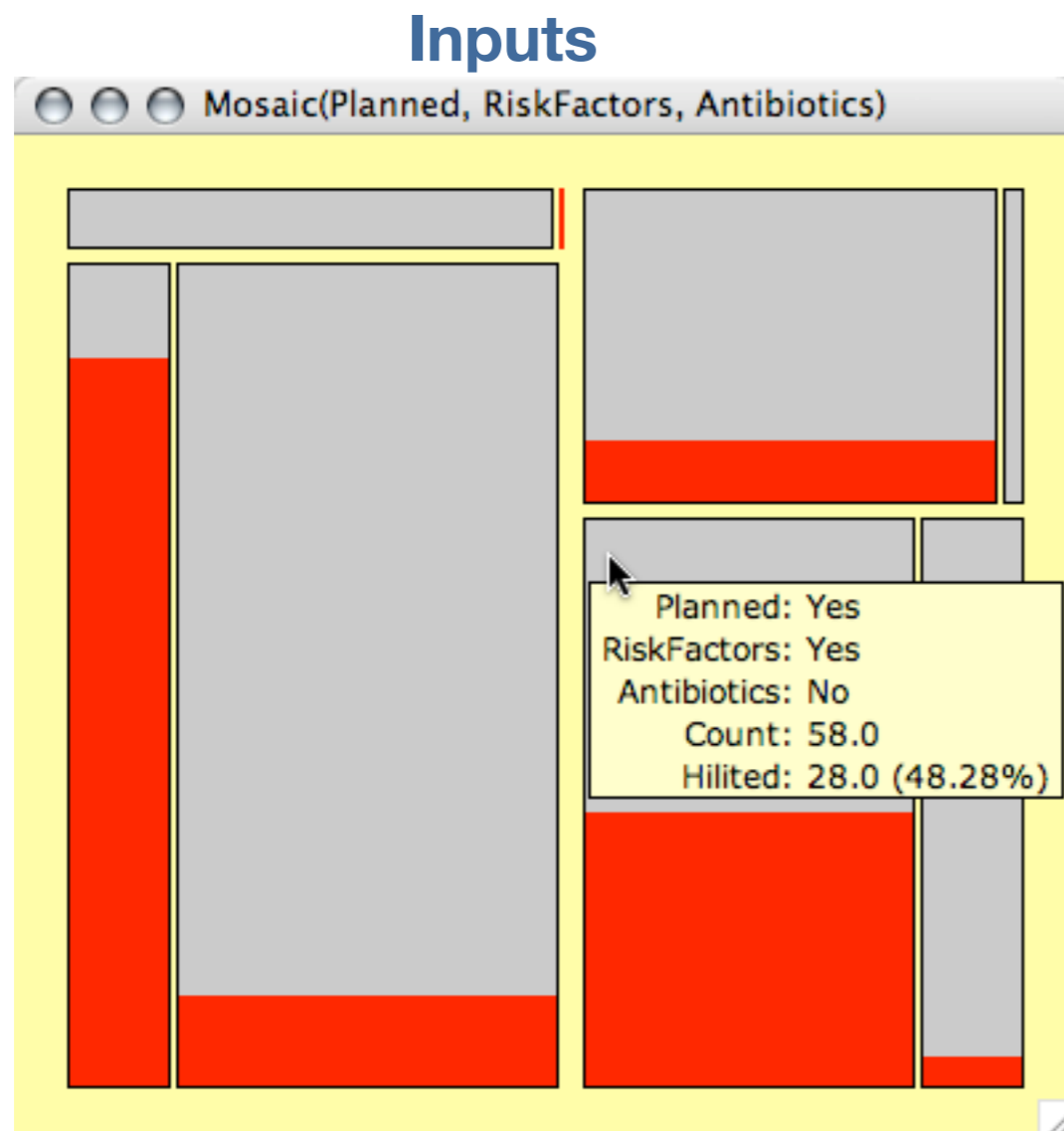
$$P(\text{Geschlecht} = \text{'male'} \mid \text{Note} < 1.25)$$

oder, die globale Schätzung mittels Dichteschätzer.

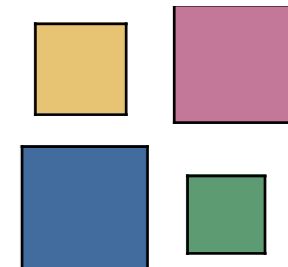


# Kategorielles Response Modell

- **Situation:**  
Multivariate kategorielle Inputs und univariaten kategoriellen Output



Gegeben, dass die Mutter zu einer Risikogruppe gehört, mit welcher Wahrscheinlichkeit kommt es zu einer Infektion?



# ANOVA

- Beispiel: Barley Data

– One way

`yield ~ species`

– Two way

`yield ~ year * site`

