

Loglinear Models: Basics

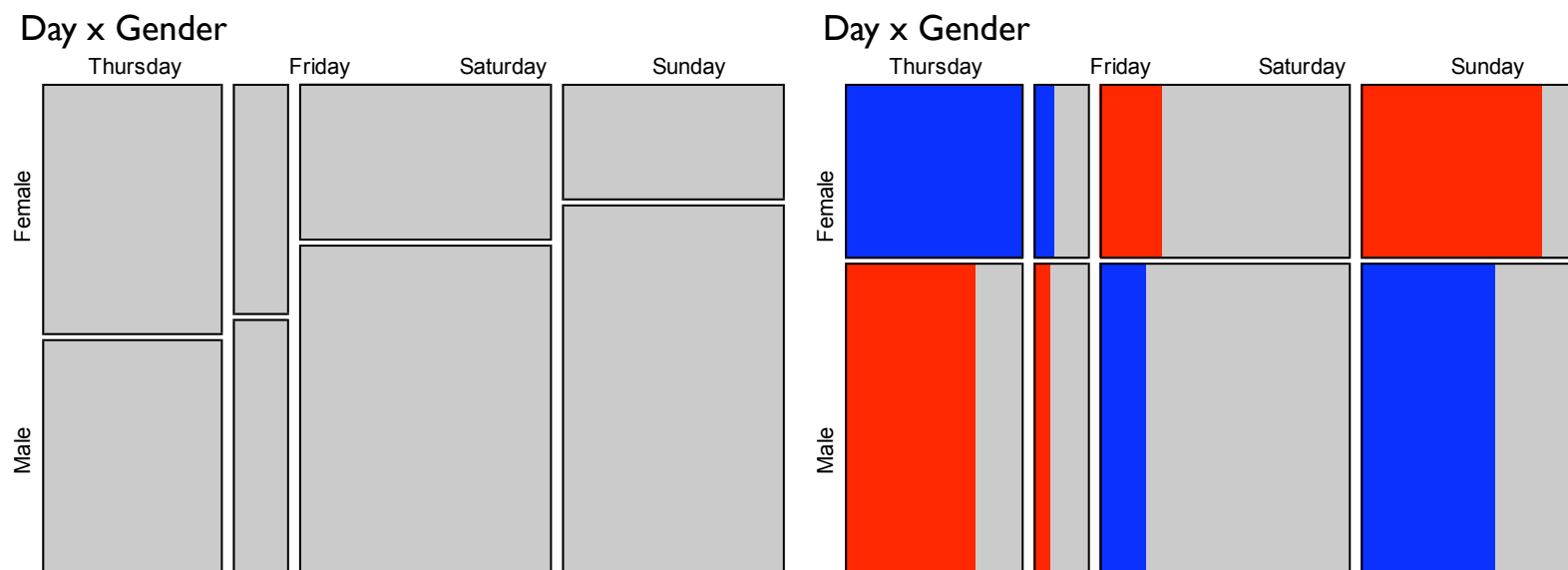
- Loglinear models can be visualized in mosaic plots just like the raw data itself
- The most trivial model would be “equal cell size”, though it is usually far from being usable for real data sets
- Marginal independence is quite easy to spot in mosaic plots, as the proportions are always the same, no matter which category we look at
- Note:
For marginal independence, the hierarchical aspect becomes less important



Loglinear Models: Visualization

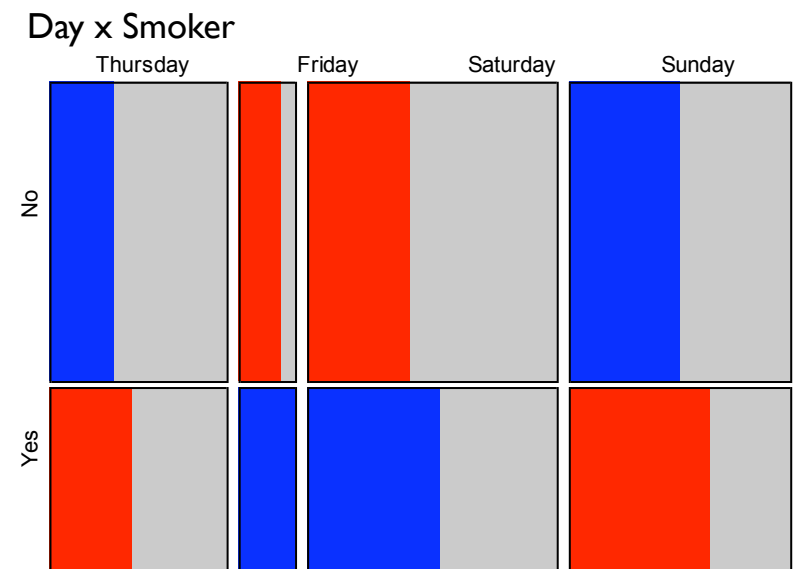
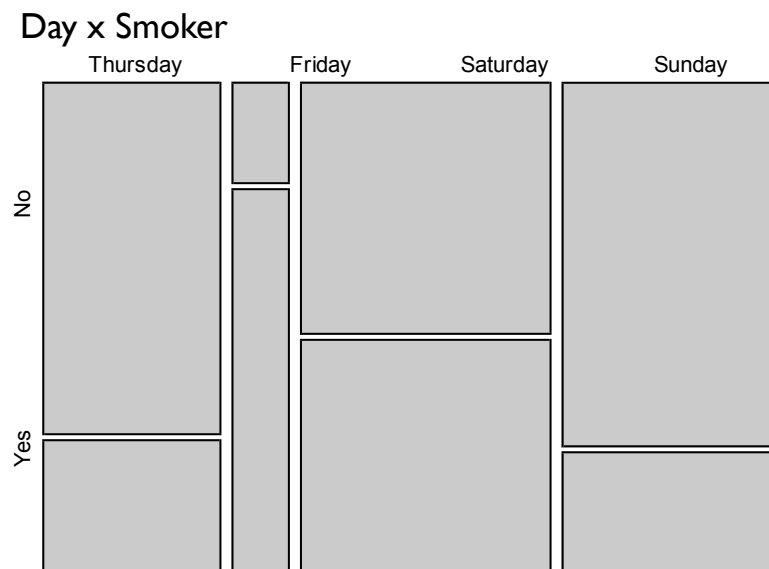
- In order to compare the deviation of the raw data to a specific model we need to code the error terms in the mosaic plot
- There are several proposals for color-coding the residuals, in Mondrian we use color for the sign, highlighting amount for the size and saturation for the overall significance of a model
- Example: Tipping Data (Appendix i)

Interaction between 'Day' and 'Gender'



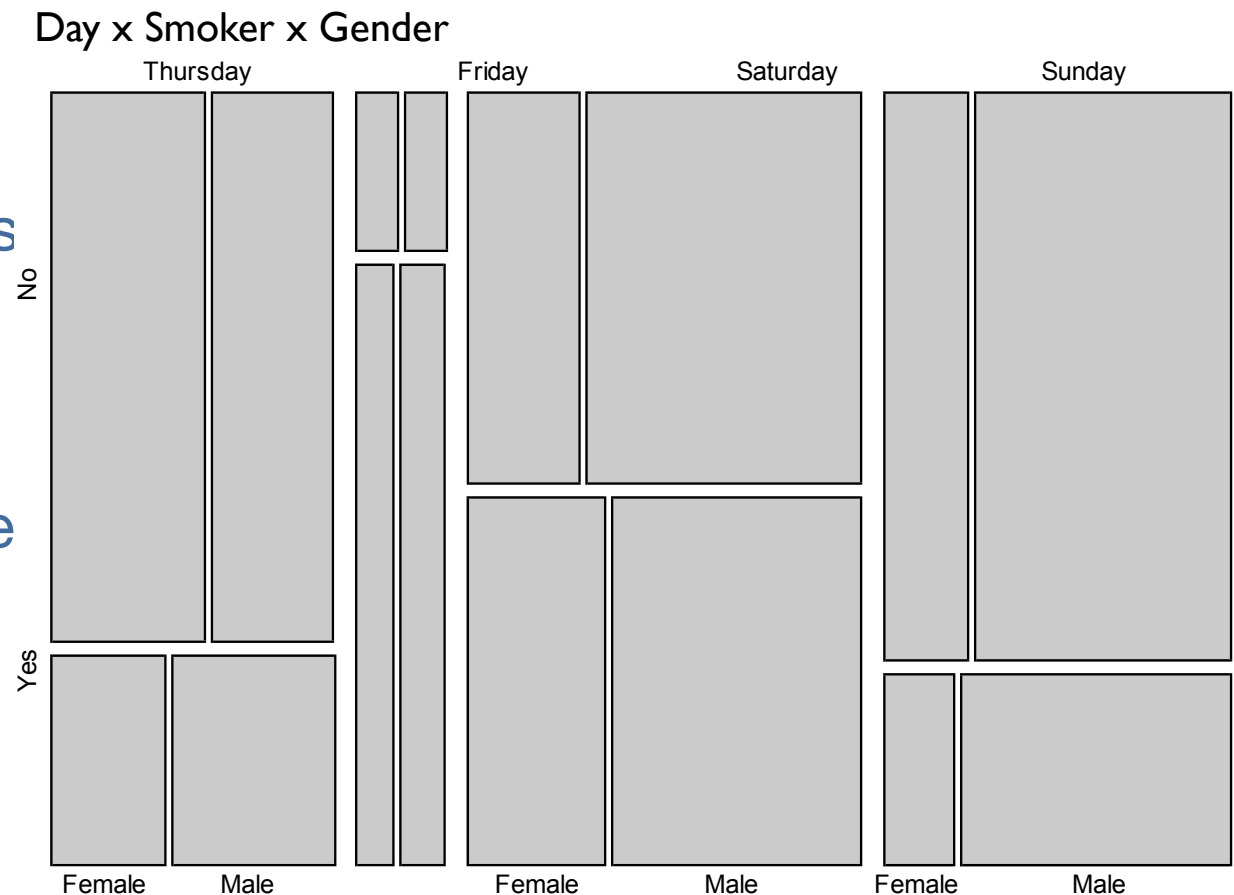
Loglinear Models: Visualization

- Whereas the interaction between ‘Day’ and ‘Gender’ had a clear trend to more males on the weekend, the interaction between ‘Day’ and ‘Smoker’ has an ‘outlier’ with far more smokers on Fridays
- In the model of marginal independence the residual highlighting is (proportionally) the largest for smokers on Fridays, although other cells have a larger absolute highlighting



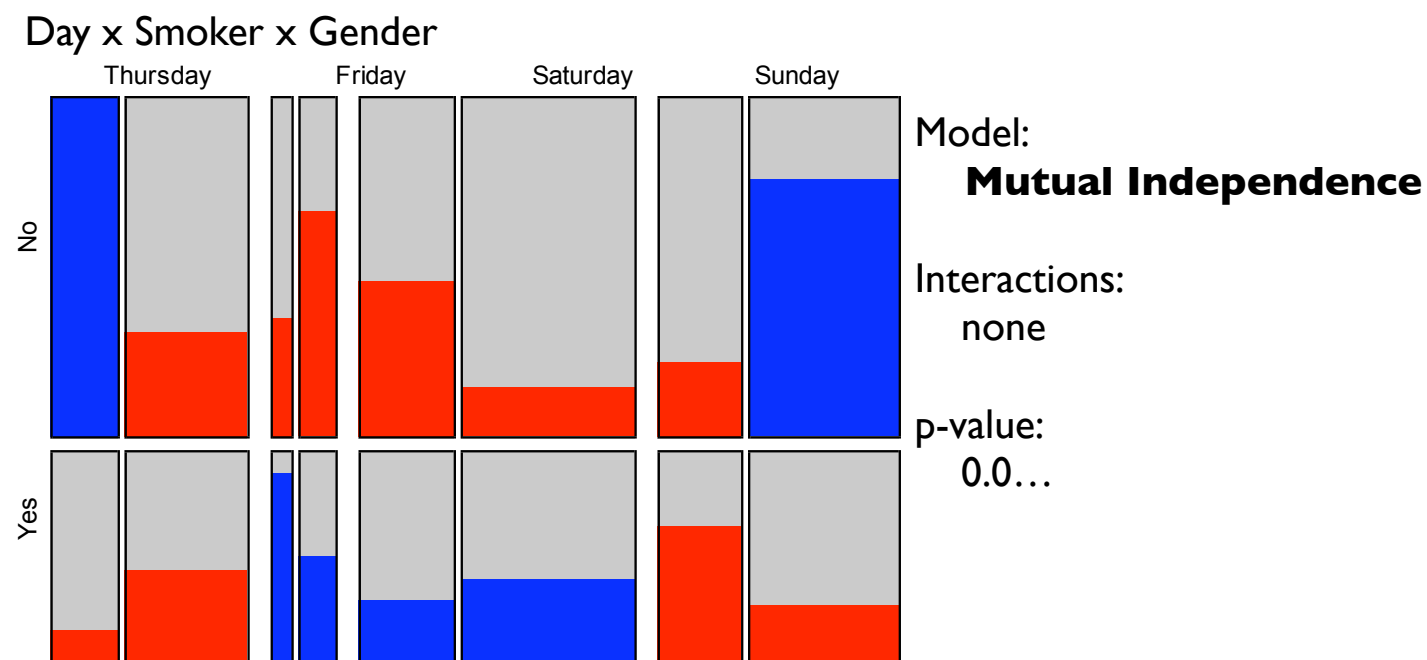
Loglinear Models: More Complex Model – Tips Data

- The mosaic plot shows the raw data for ‘Day x Smoker x Gender’
- We find the independence for ‘Smoker x Gender’ for each level of ‘Day’, though the proportions differ in the different levels
- For the three variables we have three 2-way interactions, only for one we found the case of independence



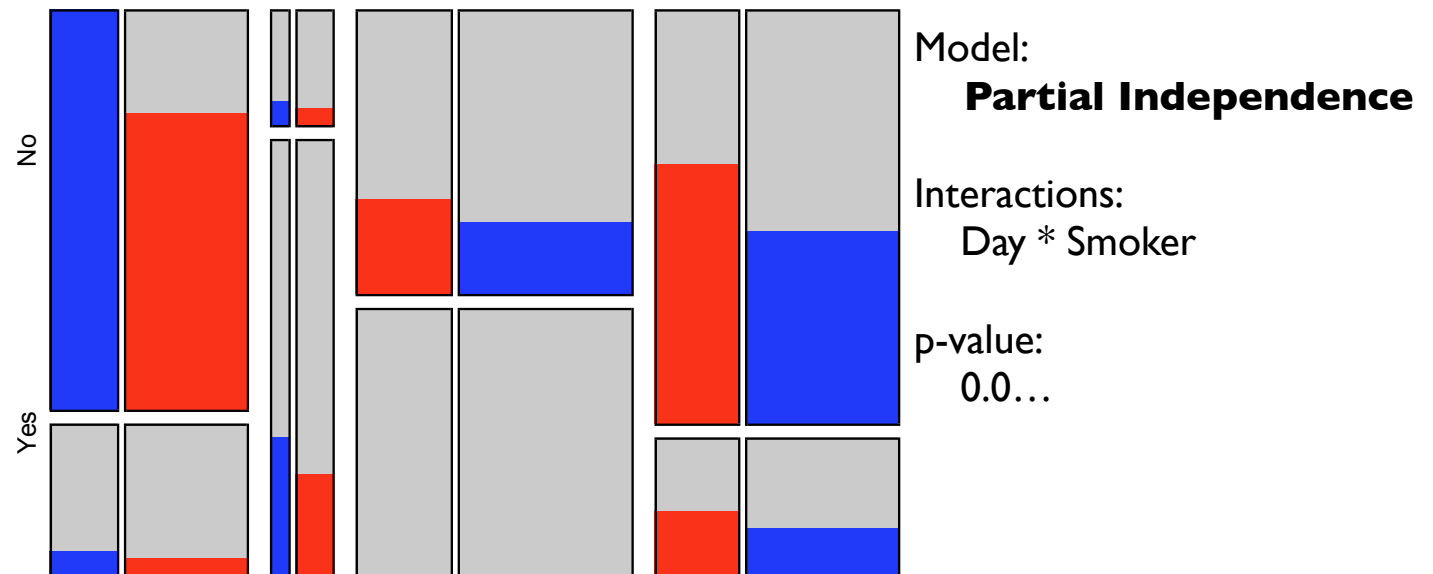
Case Study: Tips Data

- We can model the different interactions step by step and observe the impact within the mosaic plot
- In the model of mutual independence, we basically see the interaction between the first two variables which we add first



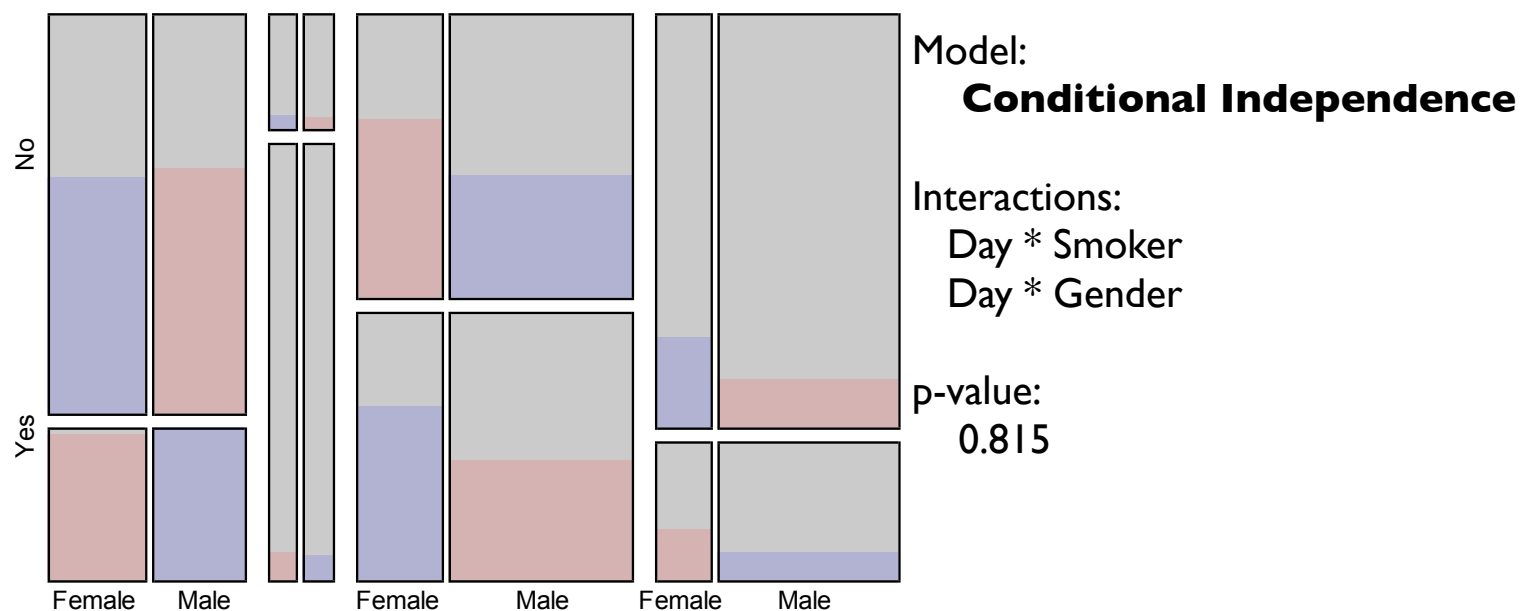
Case Study: Tips Data

- In the model plot which holds the interaction between ‘Day’ and ‘Smoker’ we see the independence for all other variables
- We find opposite colors in the highlighting for ‘Thursday’/‘Friday’ and ‘Saturday’/‘Sunday’, which indicates that ‘Gender’ and ‘Day’ may not be independent



Case Study: Tips Data

- In the final plot, the overall shape of the model plot does not differ much from the one of the raw data
- The low saturation of the residual highlighting indicates that the model is no longer significant



Loglinear Models: “Champions League”

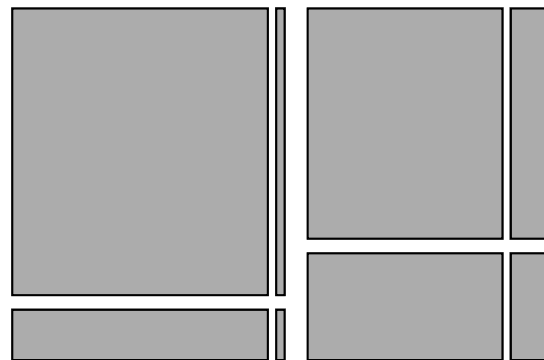
- Although it is not the usual case to look at “anonymous” data, we can assign the different loglinear models purely from the shape of a mosaic plot – which is which

1. the raw data

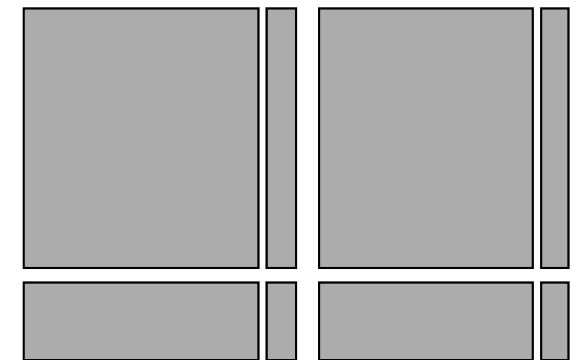
2. marginal independ.

3. partial independence

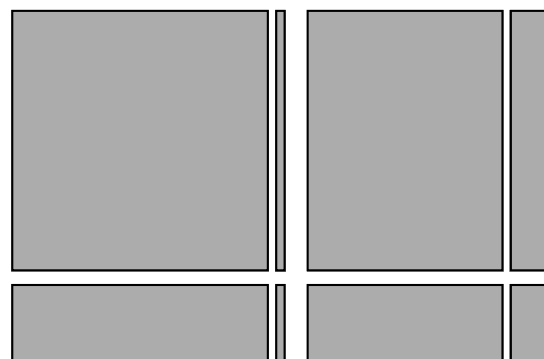
4. conditional independ.



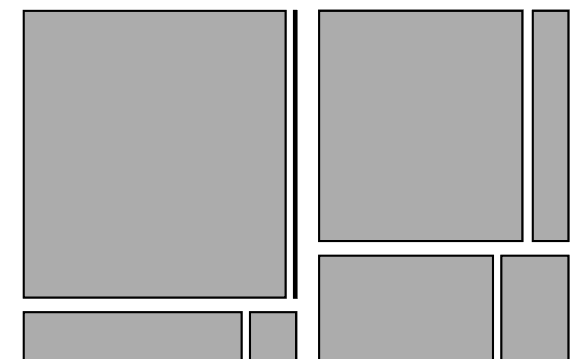
a)



b)



c)



d)