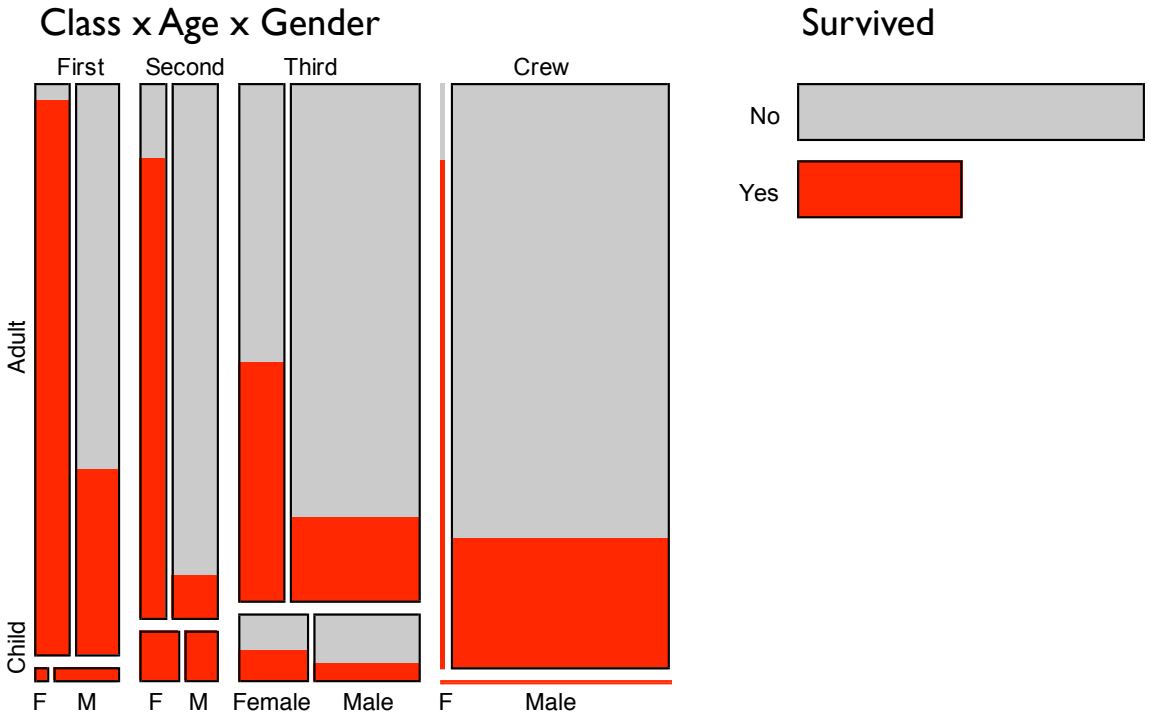
Chapter 5

Plot Ensembles and Statistical Models

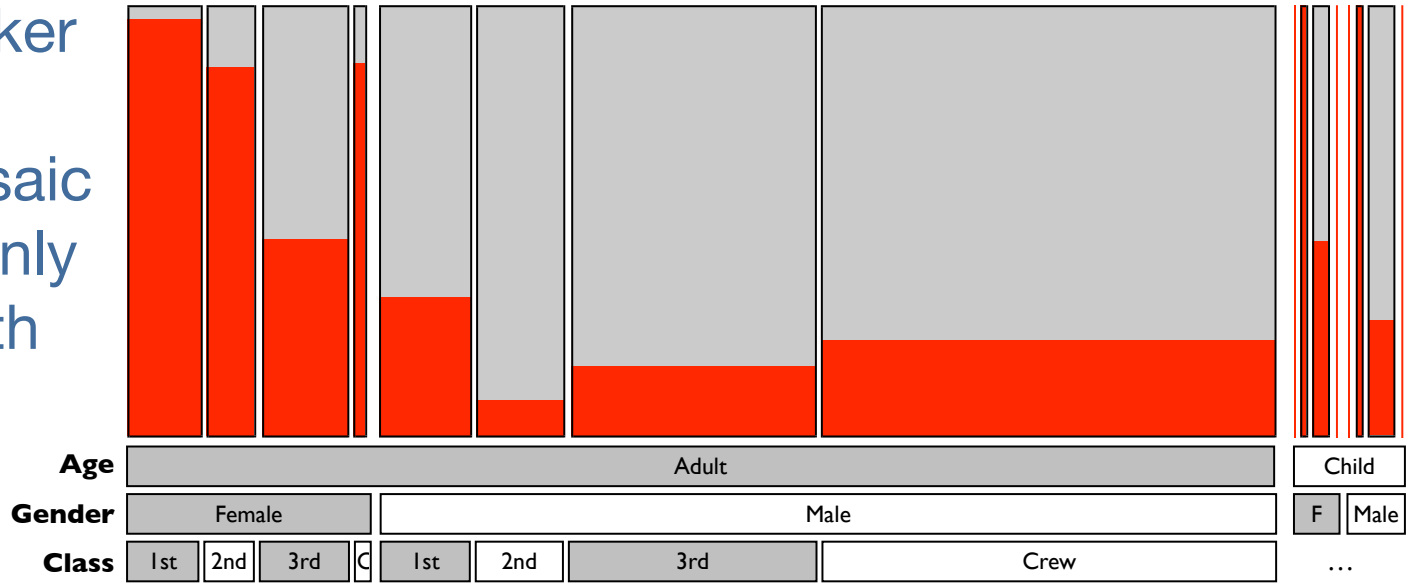
Response Models: Categorical Inputs

- In statistics we are used to think in terms of models, i.e., we think of a formal (usually functional) relationship of input variables and dependent variables and their distributions.
- For the different model classes there are certain plot ensembles which make the selection and the assessment of the inputs more transparent.
- For many categorical inputs and a categorical response a mosaic plot is the best choice.



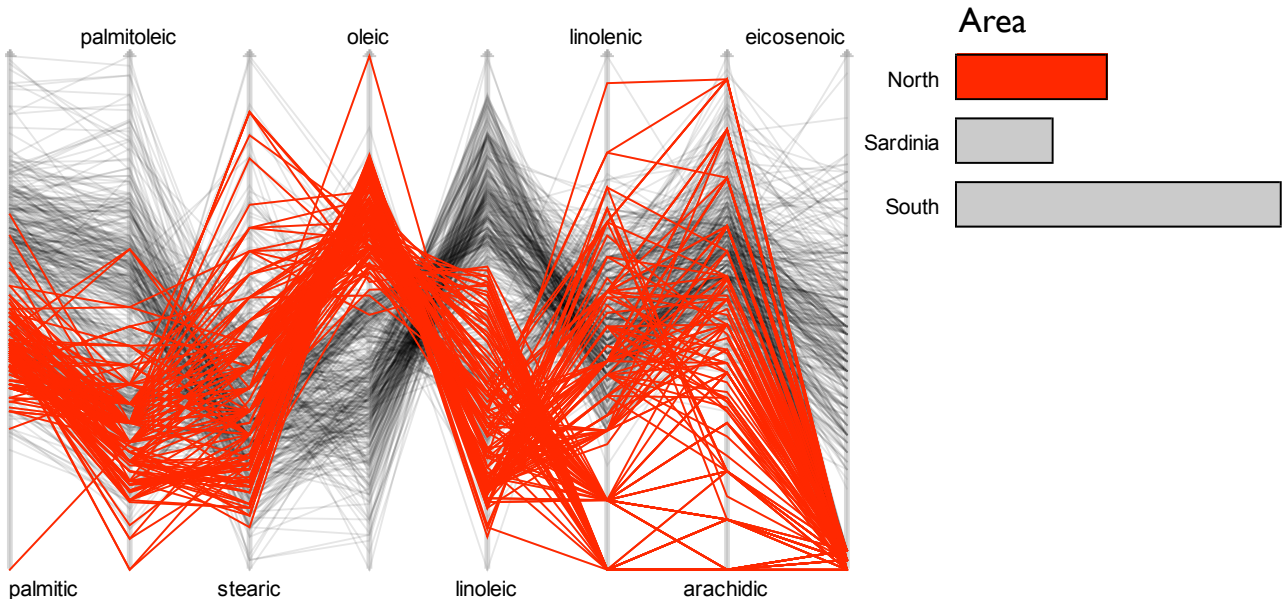
Response Models: Categorical Inputs

- **Double Decker Plots**
  In order to optimize the comparison of all crossings of all input variables, an double decker plot is most efficient.
- In a Double Decker plot the conditional layout of the mosaic plot is linearized by putting all category crossings side by side just as in a spine plot.
- The Double Decker Plots is nothing more than a mosaic plot with splits only along “x” but with additional labels



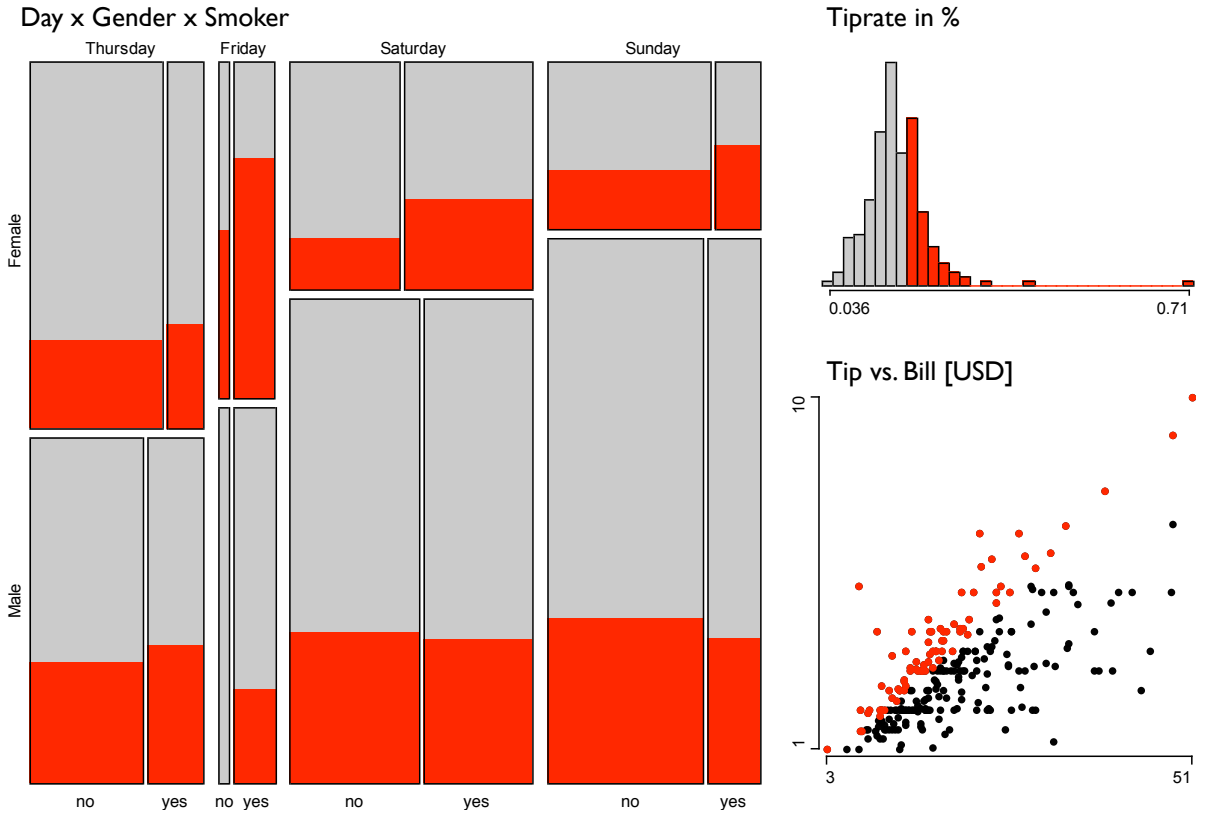
Response Models: Continuous Inputs

- With purely continuous inputs a parallel coordinates plot is the best tool to review all variables at once
- This plot ensemble (PCP + barchart) makes it easy to rule out variables which either do not separate at all or are completely unusable for a certain kind of model.
- We can not easily judge from the highlighting in a PCP how efficient a certain variable will be



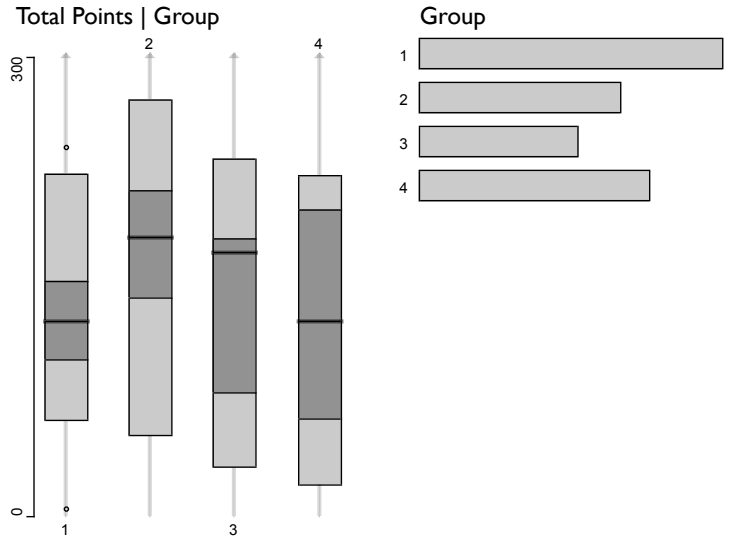
Response Models: Mixed Inputs

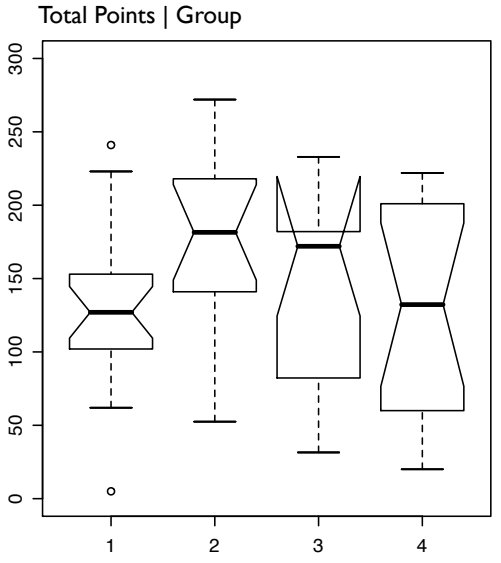
- For a set-up which has a mixture of continuous and categorical variables, we also need a mixture of plots which hold the input variables.
- Linking the different plots still gives us the possibility to look at all inputs at a time
- To investigate inter-action structures we need more complex selections

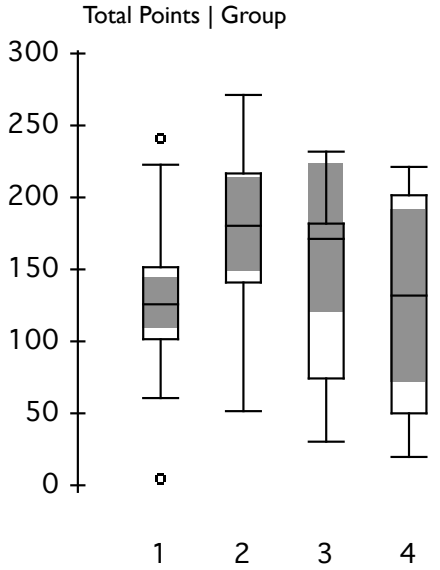


ANOVA: Standard Graphics (1-way)

- Classical ANOVA is build upon means and variances
- Nonetheless, regarding a comparison of the distributions, boxplots are more robust than means and variances
- Although the size of a subgroup is already indicated by the corresponding variance, using barcharts alongside helps
- Notched boxplots or other modifications may facilitate an ad-hoc comparison of medians

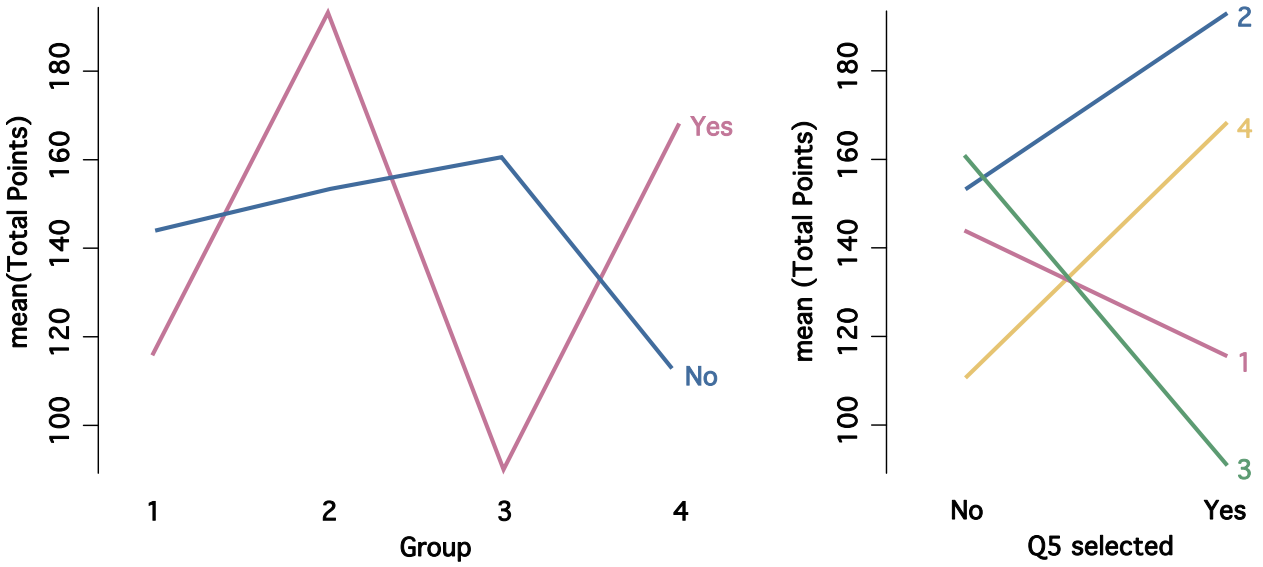






ANOVA: Interaction Plots (2-way)

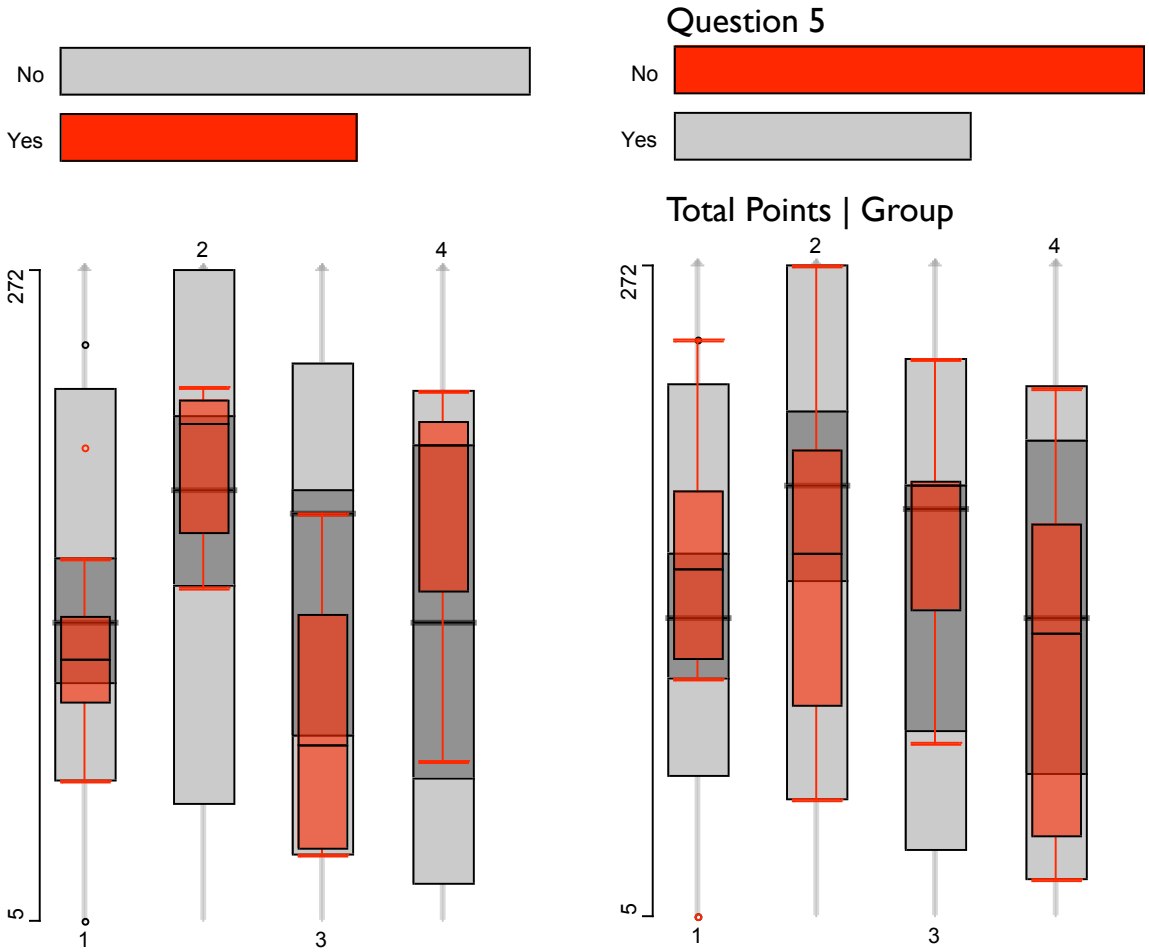
- 1-way ANOVA – i.e., a simple comparison of means – is usually not enough when we want to comprehensively look at data
- The comparison of medians along one scale is efficient, a 2-dim trellis-like set-up makes the comparison of medians much harder
- The so called interaction plot traces group means of one factor across all levels of another factor
- Example:
  Interaction between "Group" and "Q5 selected" from the Exams dataset in case study A



ANOVA: Detecting 2-way Interactions I

- In an interactive environment we may switch the selection between groups of one factor and observe the highlighting of the other factor
- Thus we do not need a new plot but only deploy an interactive technique
- Example:
  Student's Exam – CS A

  "Is there an interaction between the study group and whether or not one chose to work on question 5?"



ANOVA: Detecting 2-way Interactions II

- The role of the two factors can be interchanged, i.e., one factor is displayed in the boxplot y by x and the other is used to select the different levels
- In either case, if no interaction is present, we expect the medians in the boxplot to change in the same direction when we switch levels in the barchart.
  If the medians change in opposite directions we see an indication for an interaction