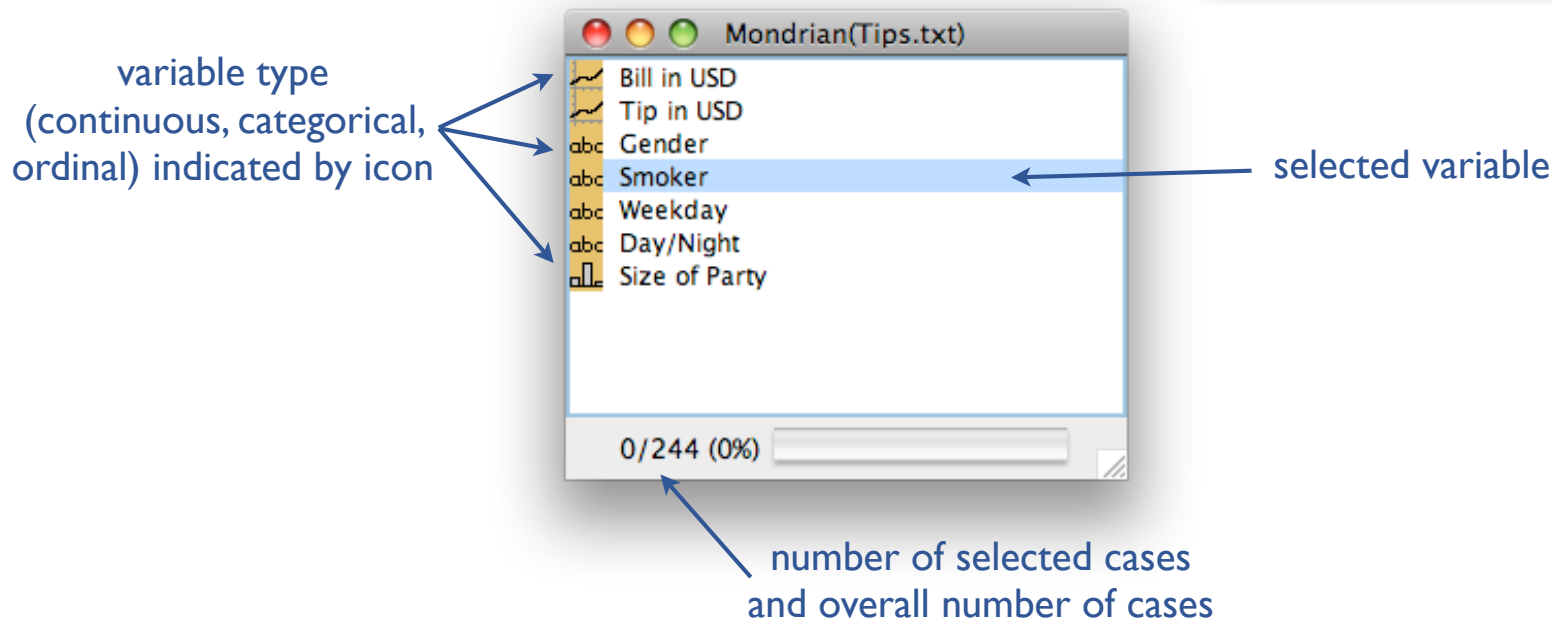
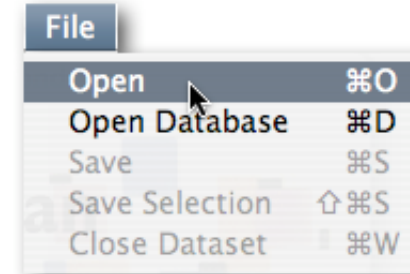


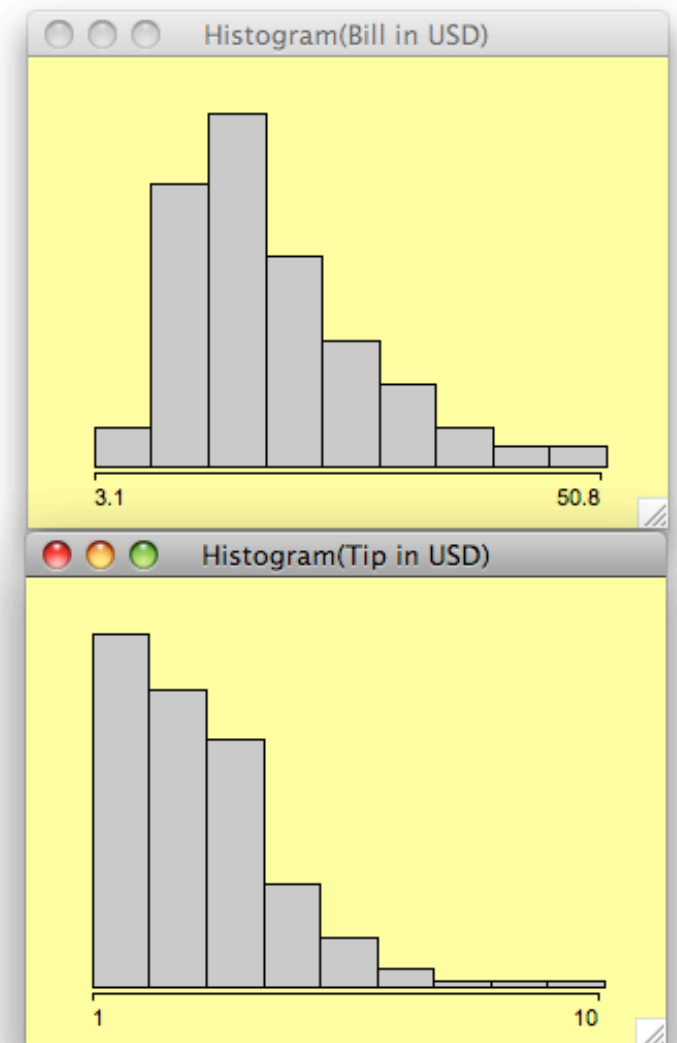
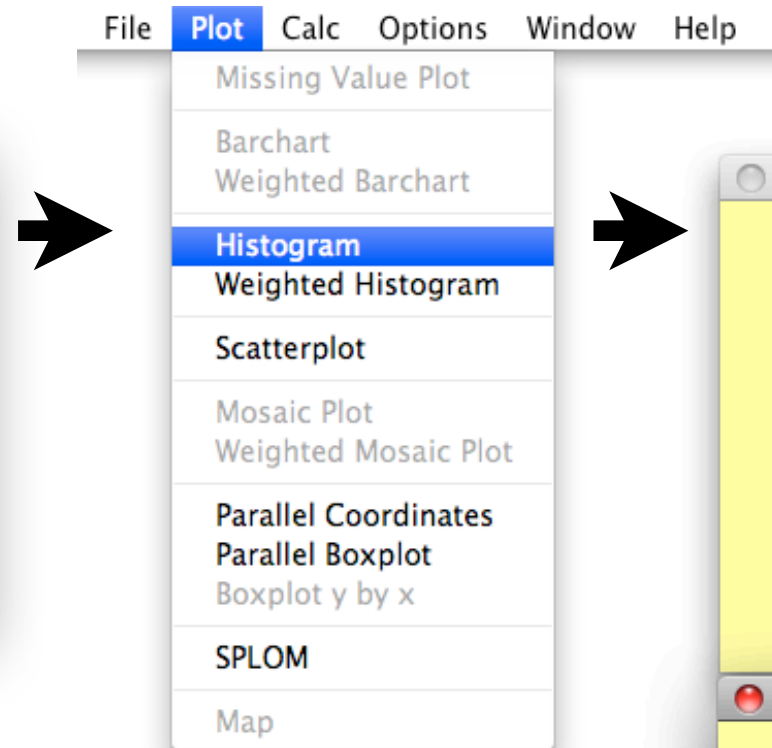
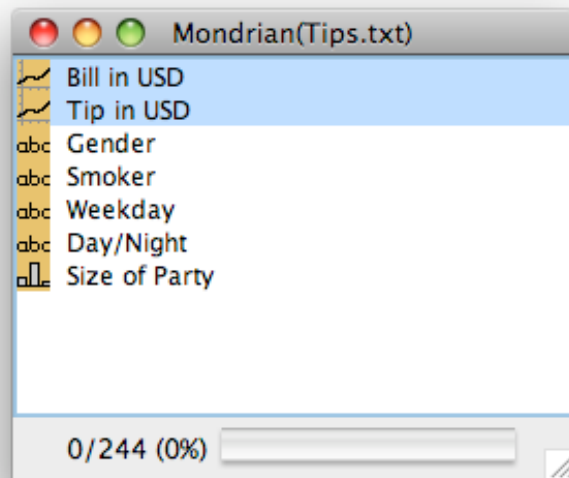
Hands on Introduction to Mondrian Tipping Data

Introduction

- Download Mondrian:
 - <http://www.rosuda.org/Mondrian>
- Start, open dataset: File → Open
- Variables window:

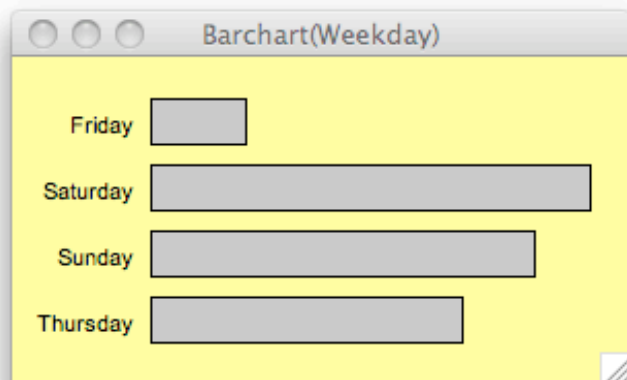
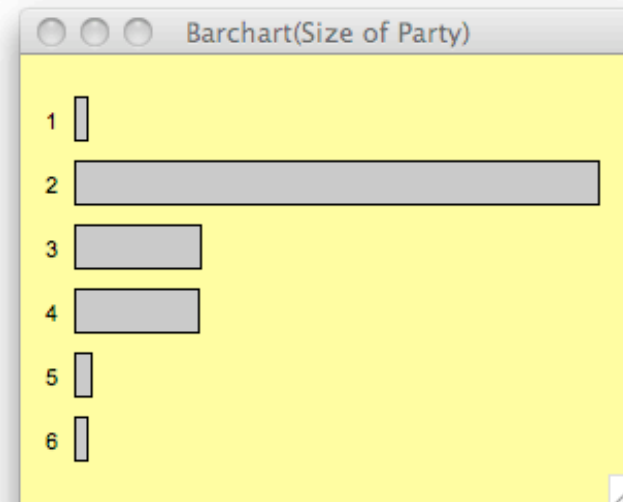
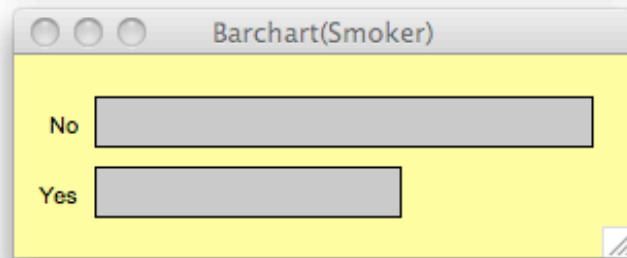
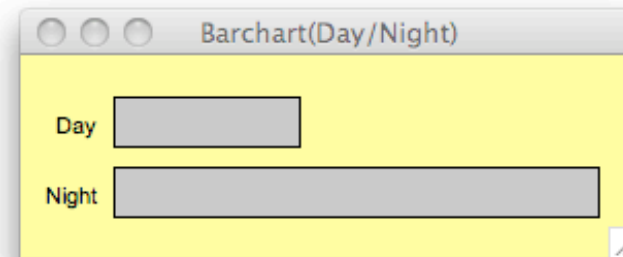
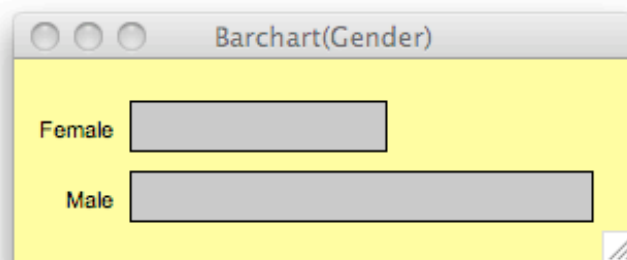


First Plots



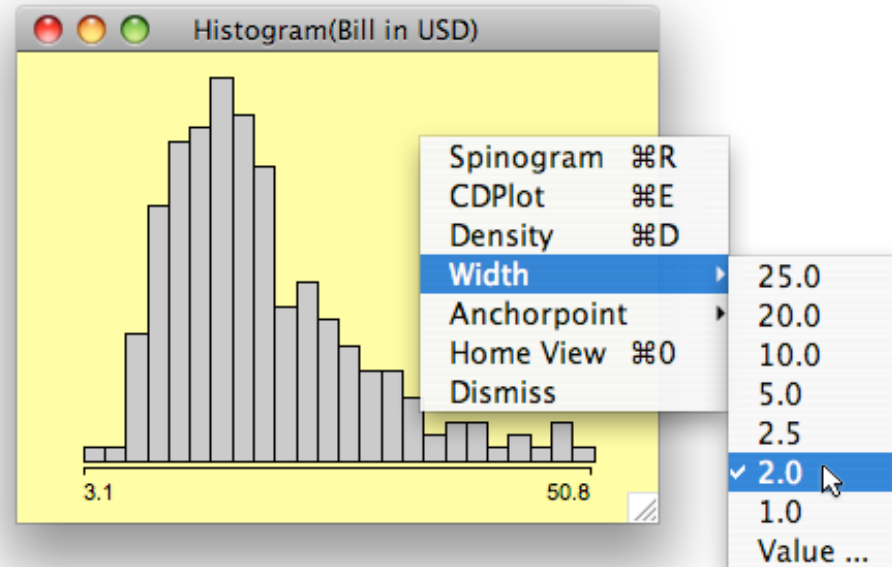
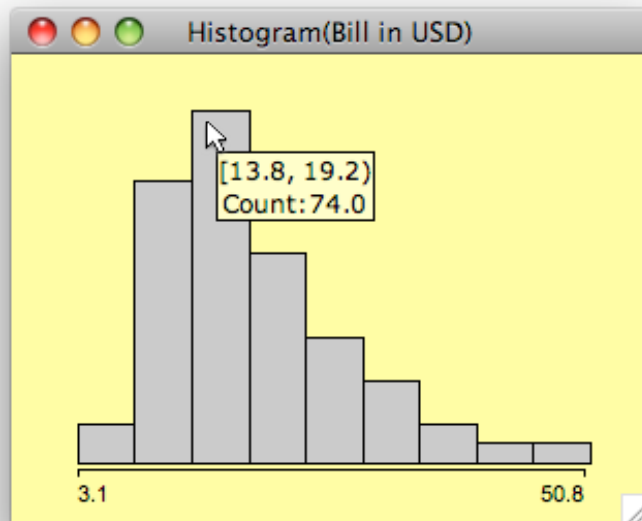
- Select variables
- Choose a plot from the Plot menu

Barcharts of categorical and ordinal variables



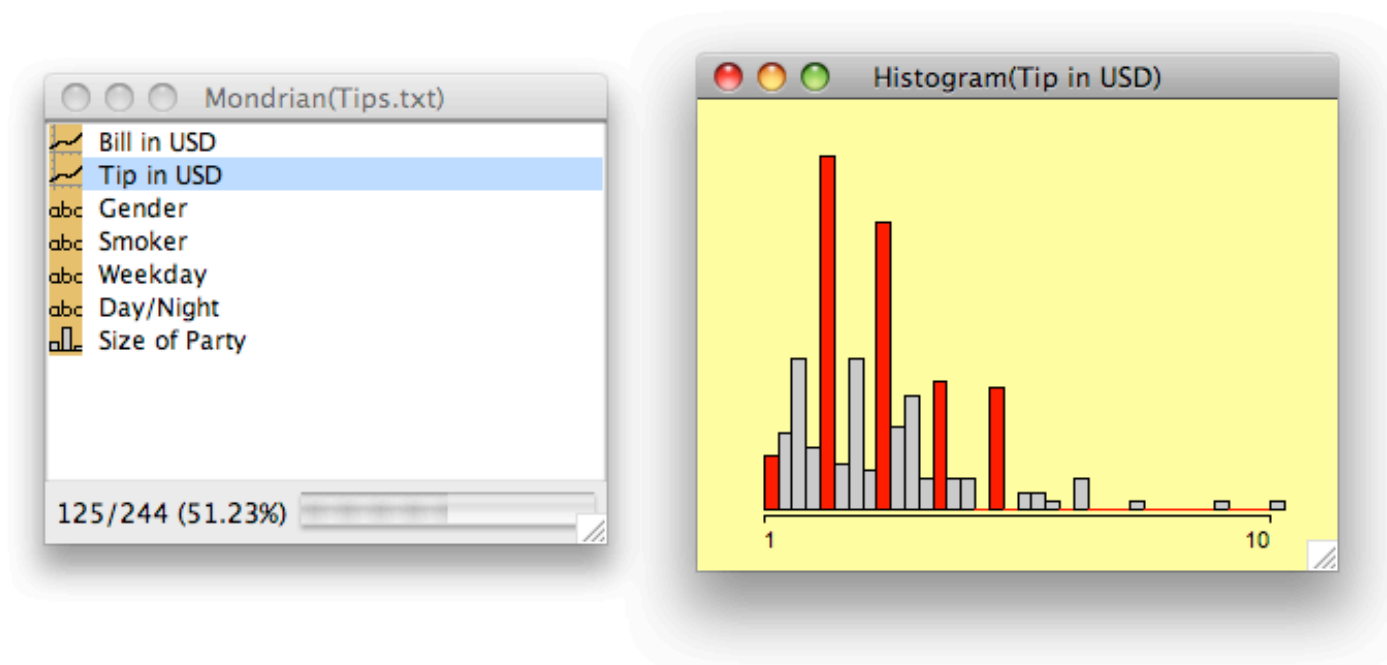
Plot Interactions

- Queries
 - display information about the element pointed by the cursor
 - hold “Ctrl” (Control) key to invoke
- Change plot parameters
 - right-click (Ctrl-click on a Mac) for context menu
 - use cursor keys



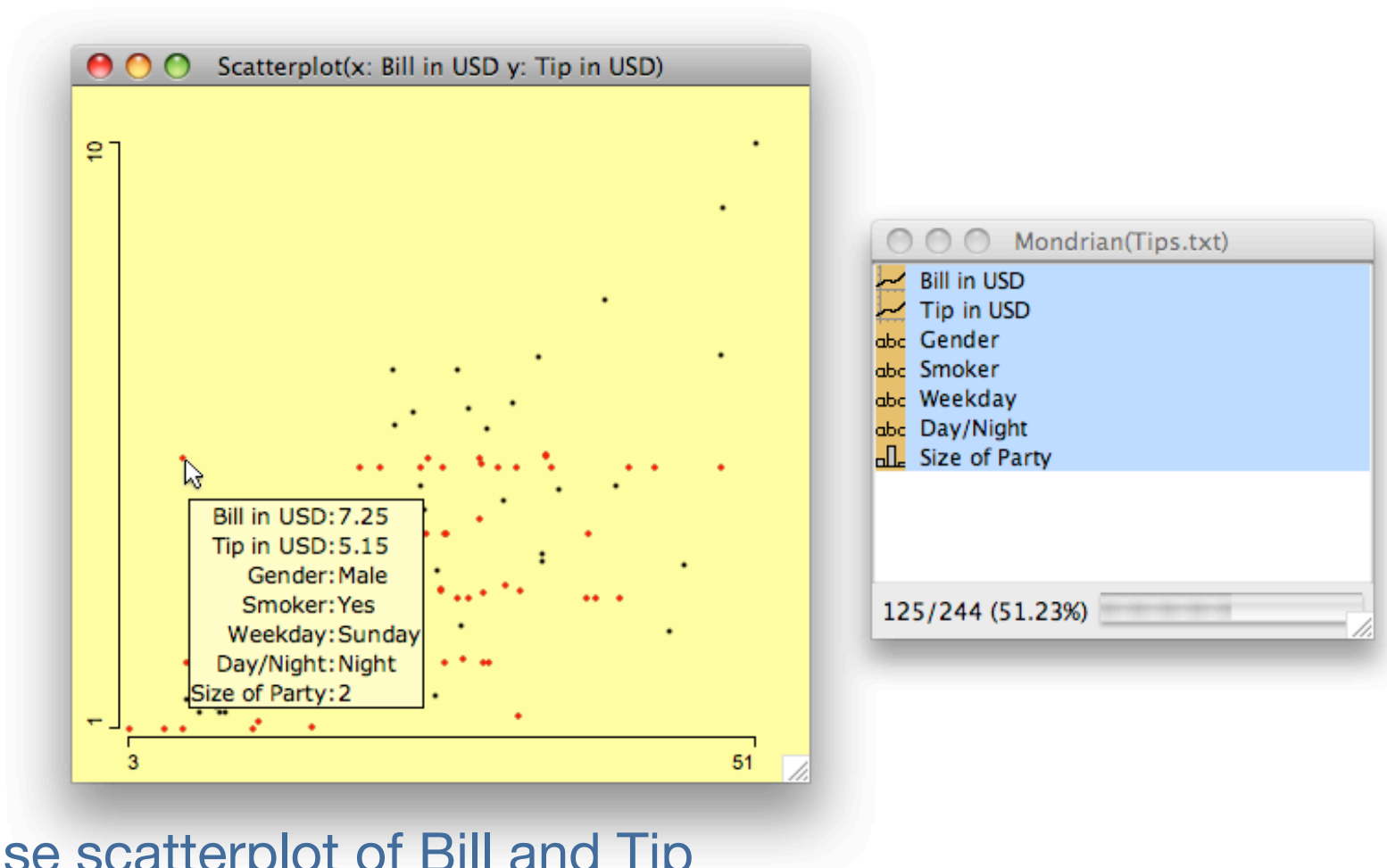
Investigating tips

- Set bin width for Tip in USD to \$0.25



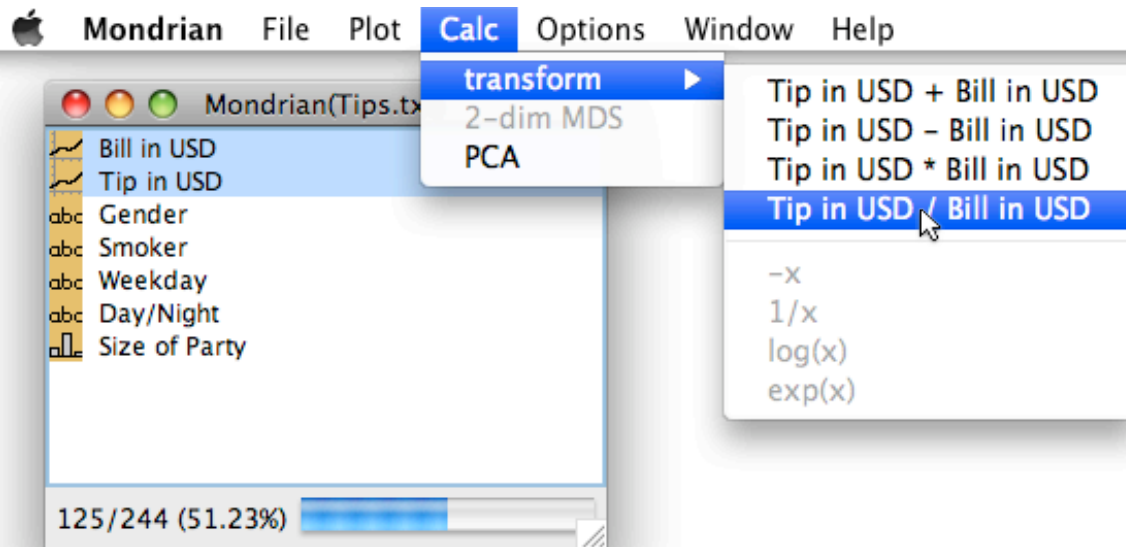
- Select bars corresponding to \$1, \$2, \$3, \$4 and \$5 by holding the <Shift> key
- More than 50% of bills are selected (see variable window)

Bill vs Tip relationship

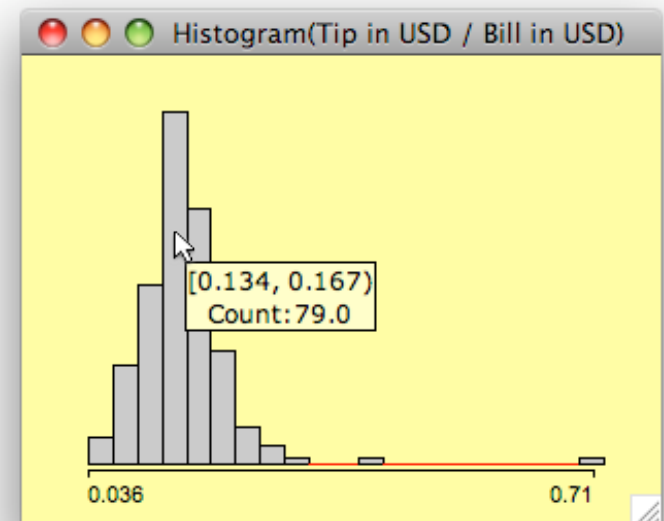


- Use scatterplot of Bill and Tip
- Select all variables
- Use extended query (<Shift><Ctrl>) for interesting outliers

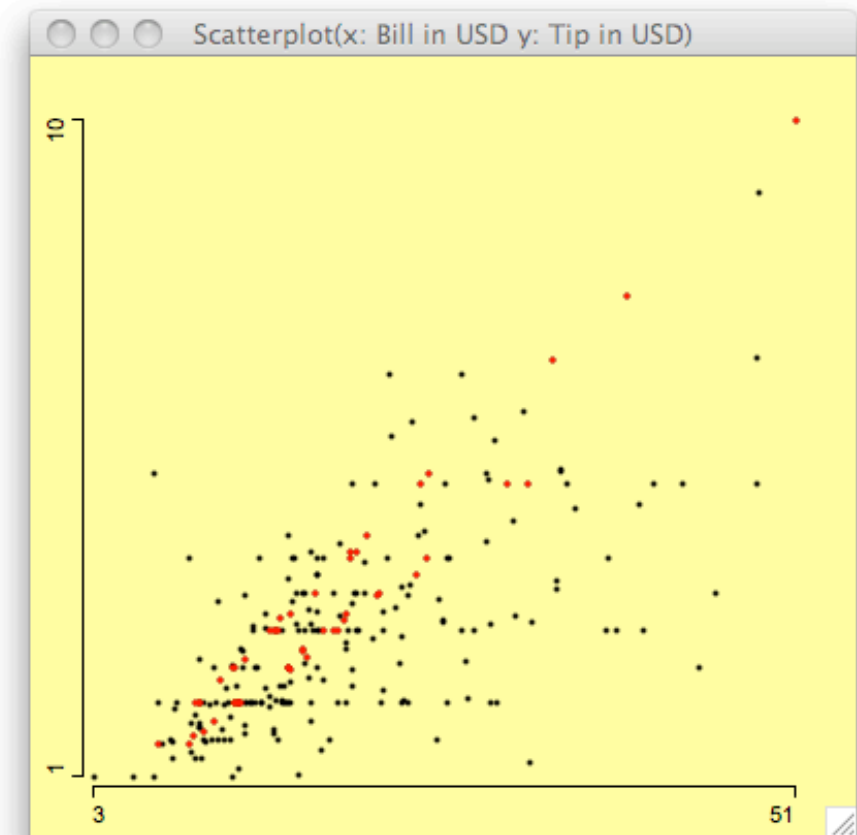
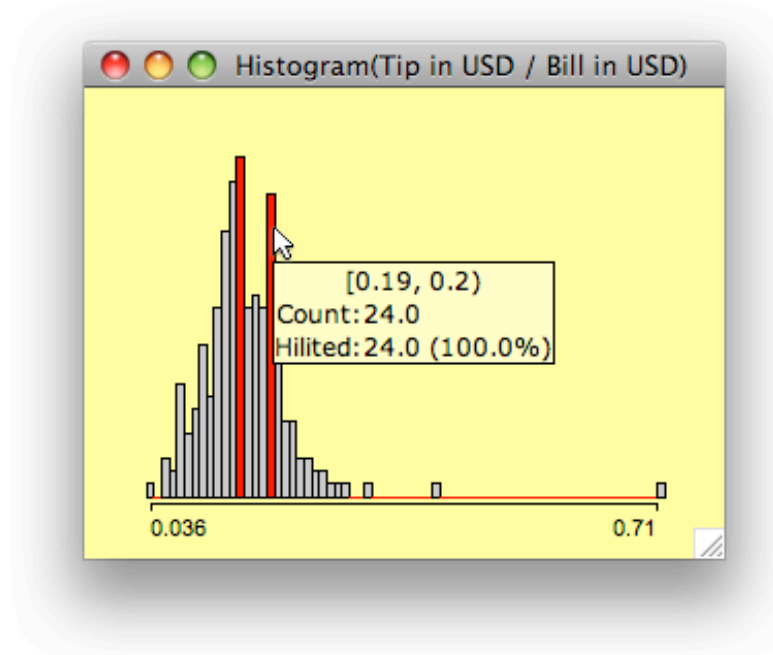
Create derived variable: Tip / Bill



- Select variables to use
 - order is important for non-commutative ops
- Select transformation in Calc
- Plot a histogram of the newly created variable - most tips are around 15%



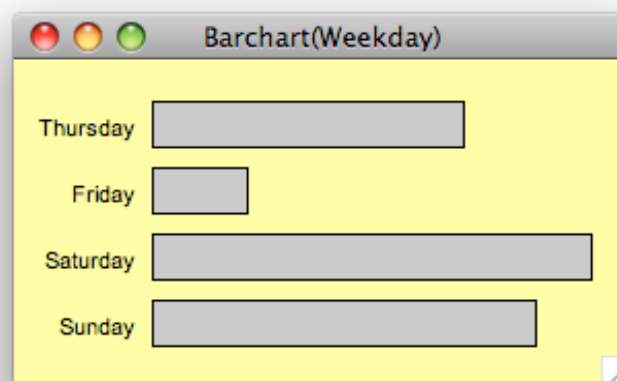
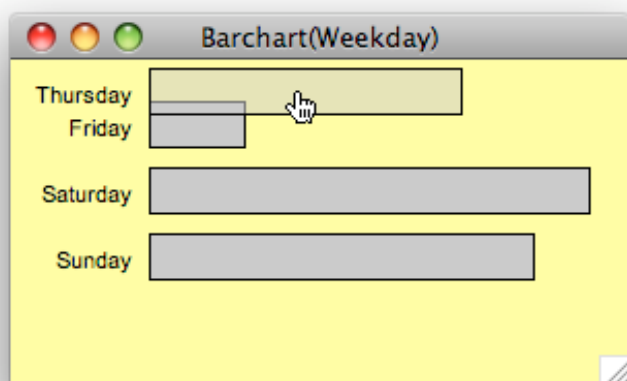
More about Tips ratio



- Set anchor to 0.0 and bin width to 0.01
- Select the two distinct modi at 15% and 19%, follow the corresponding linear relationship in the scatterplot
- Tips over 20% are rare

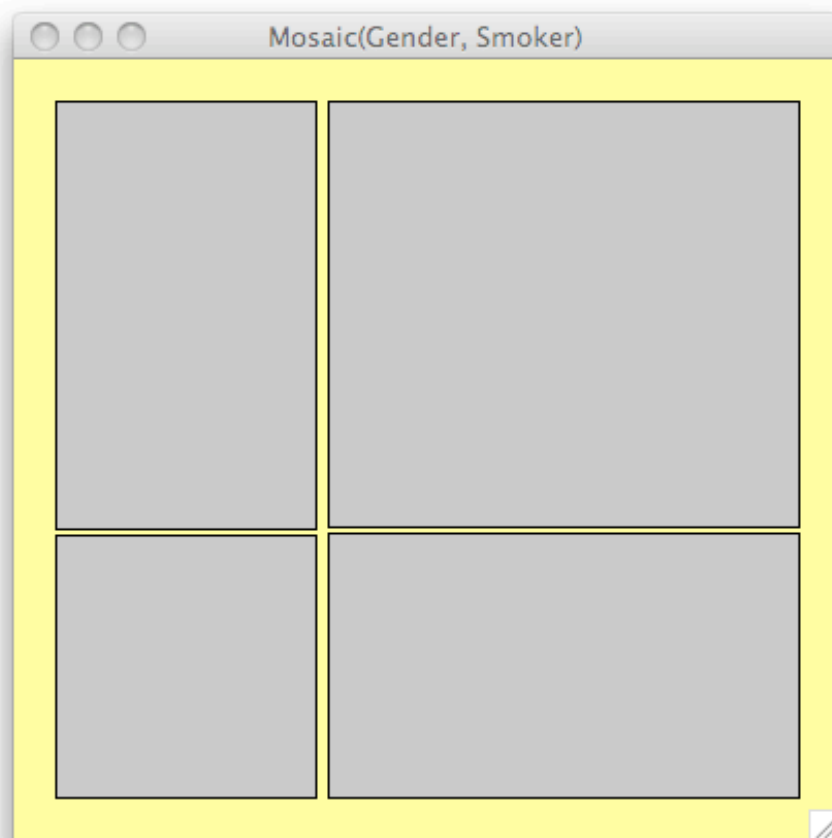
Categorical variables - order

- Re-order weekdays by holding <Alt> and dragging

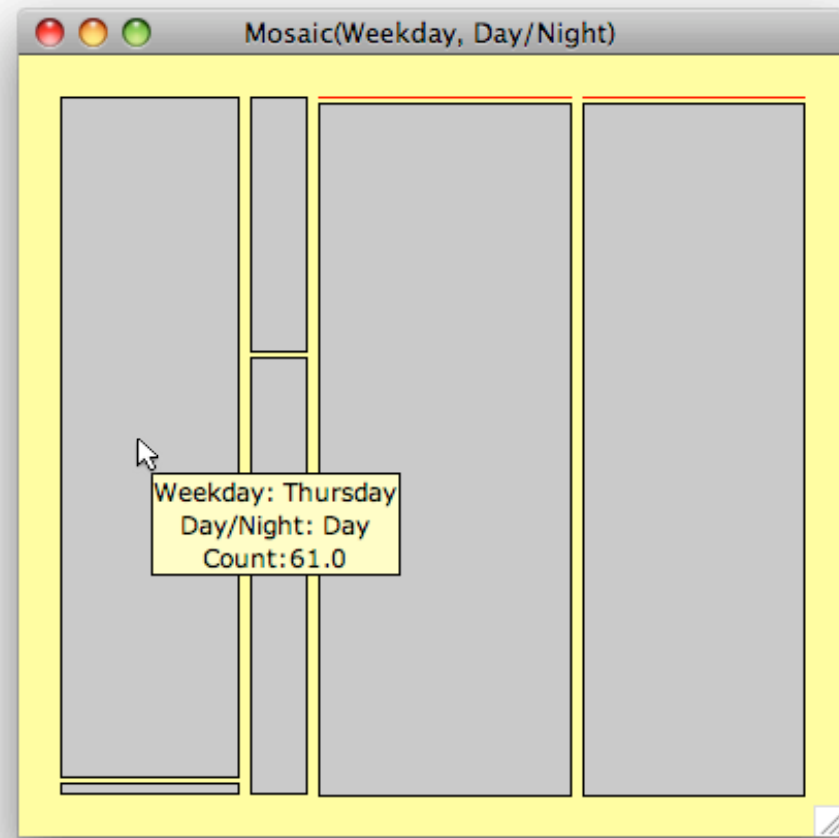


Categorical variables - relationship

- Mosaic plots



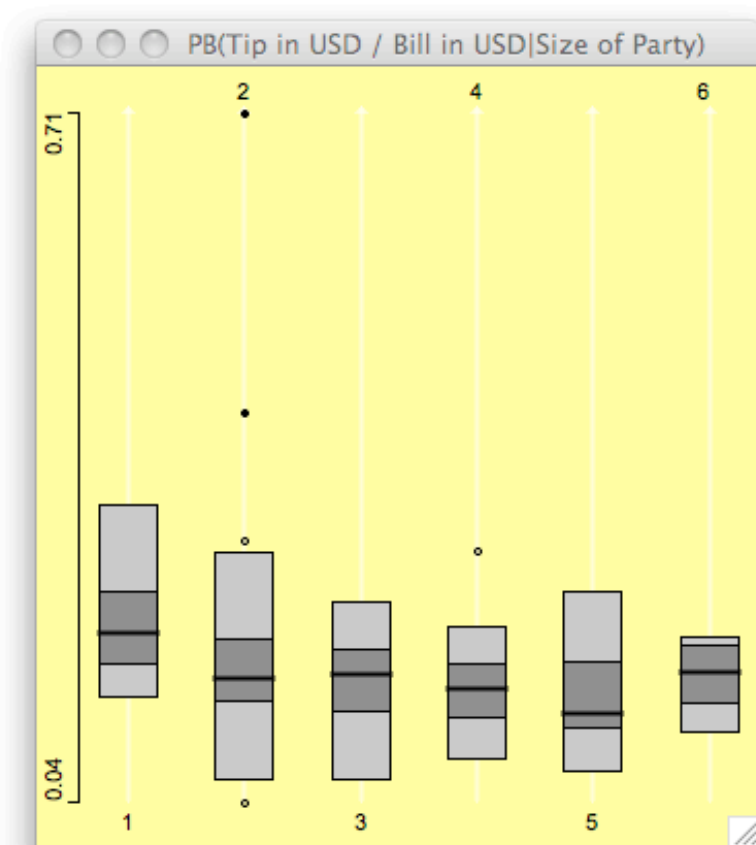
Gender × Smoker = independent



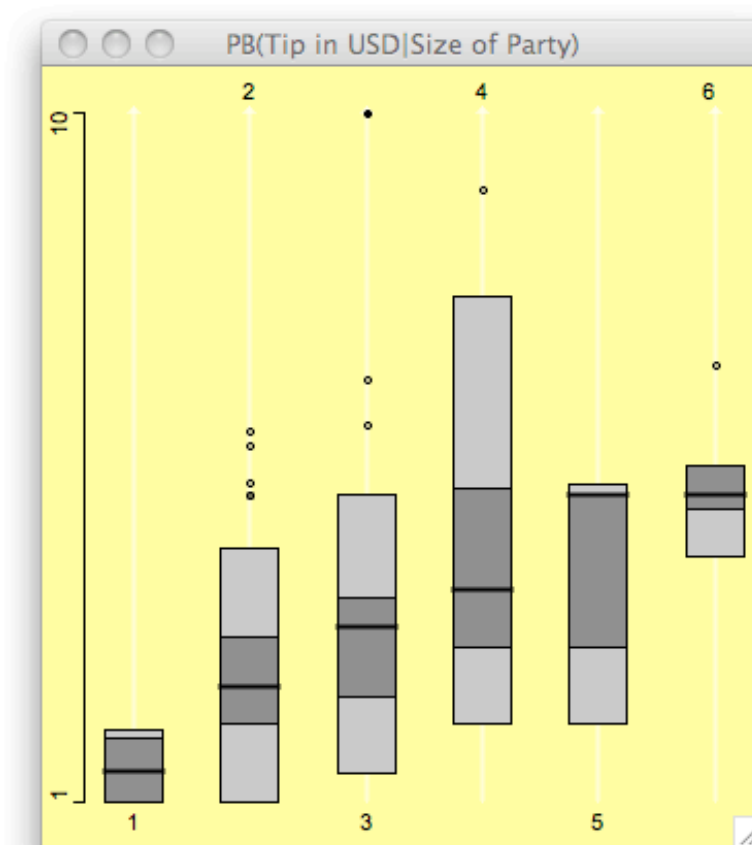
Weekday × Day/Night = strongly dependent (mostly daytime on Thursday, night on weekends)

Tip by group size

- Use Boxplots y by x

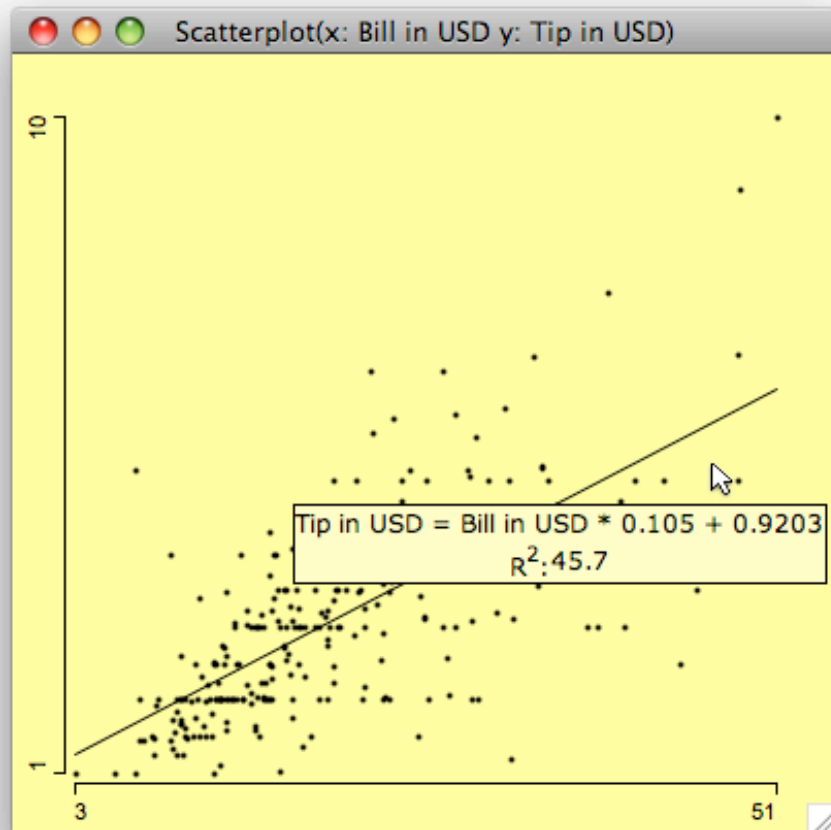


Tip ratio by group size

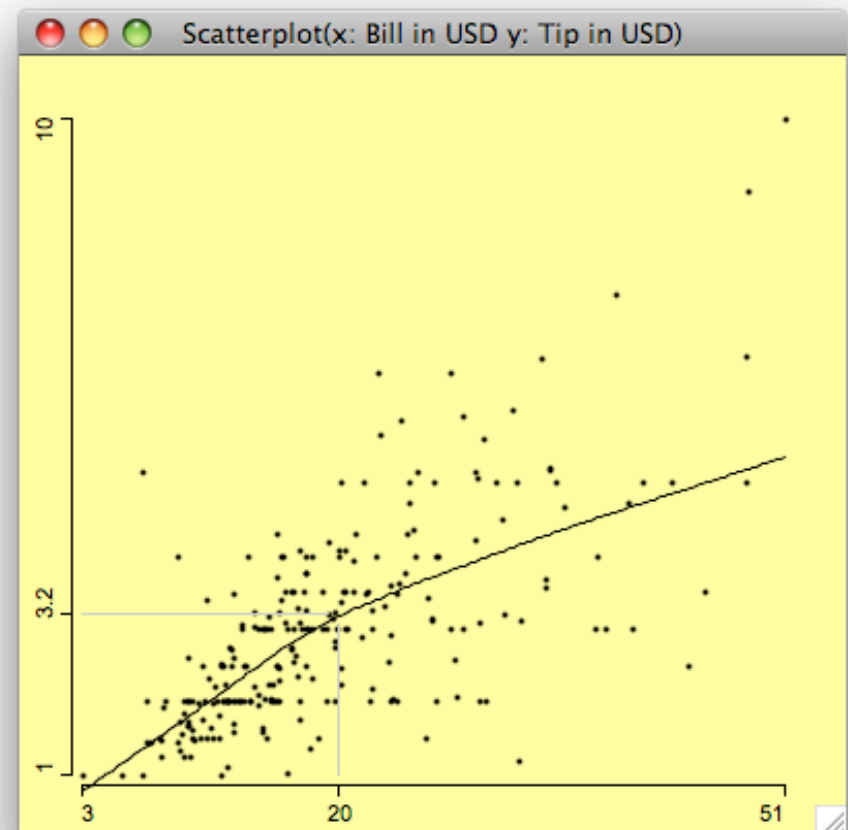


Tip amount by group size

Functional relationship

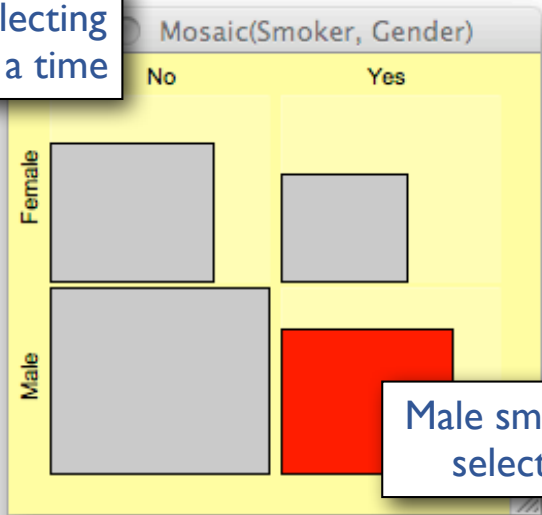


linear model

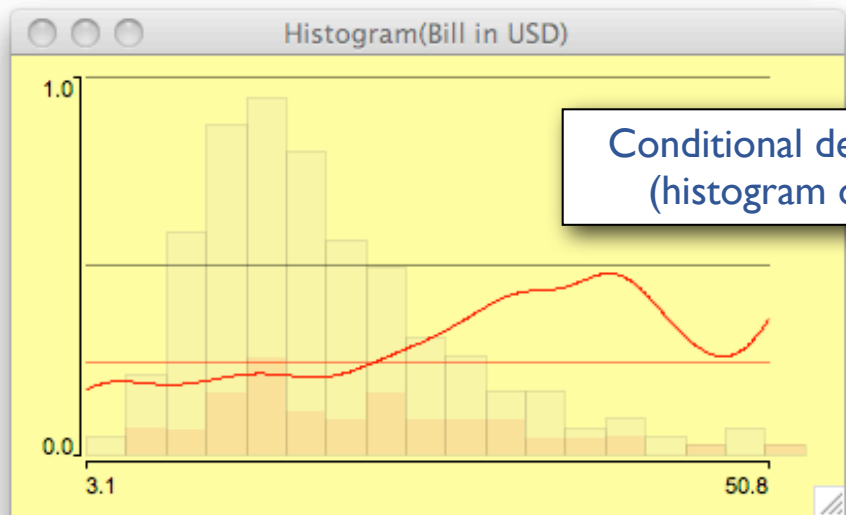


loess smoother
(note slope change at ca. \$20)

Use for selecting a group at a time

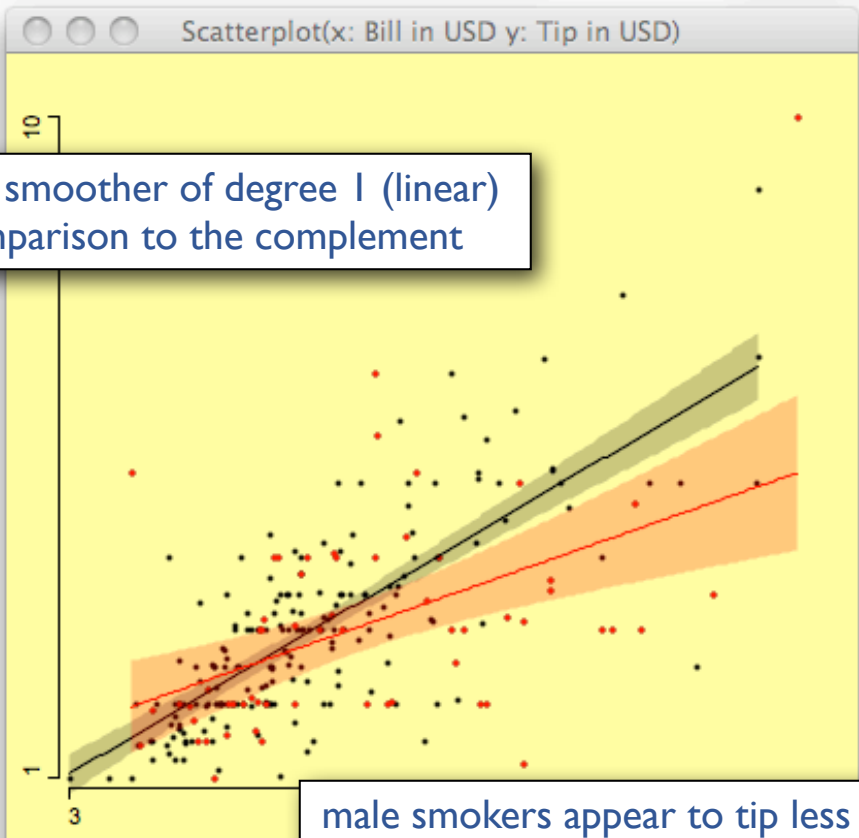


Male smokers selected



Conditional density plot (histogram overlay)

Polynomial smoother of degree 1 (linear) with comparison to the complement



male smokers appear to tip less on larger bills than others

