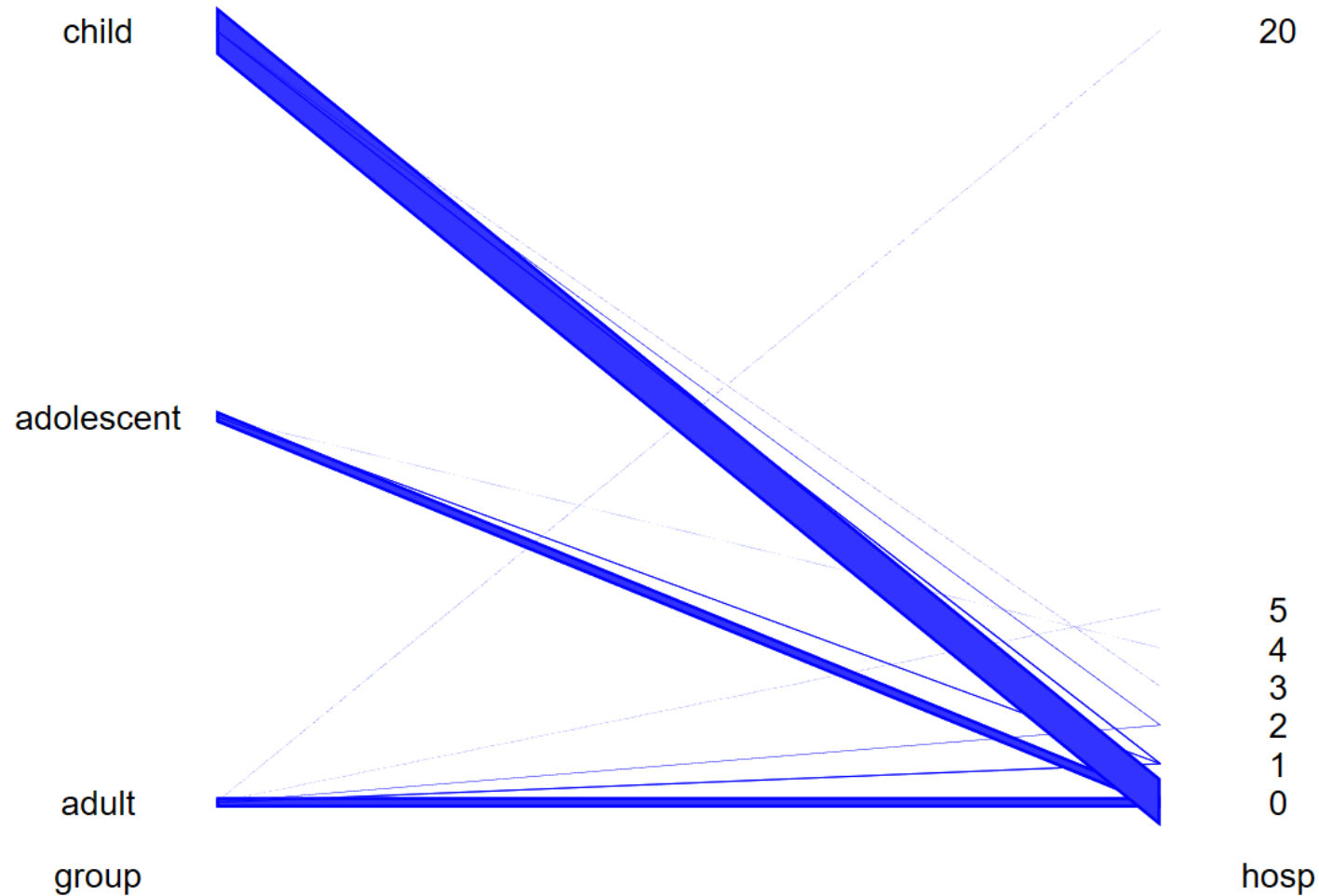


# Hammock Plots

Matthias Schonlau  
University of Waterloo

# Hammock plot with 2 variables

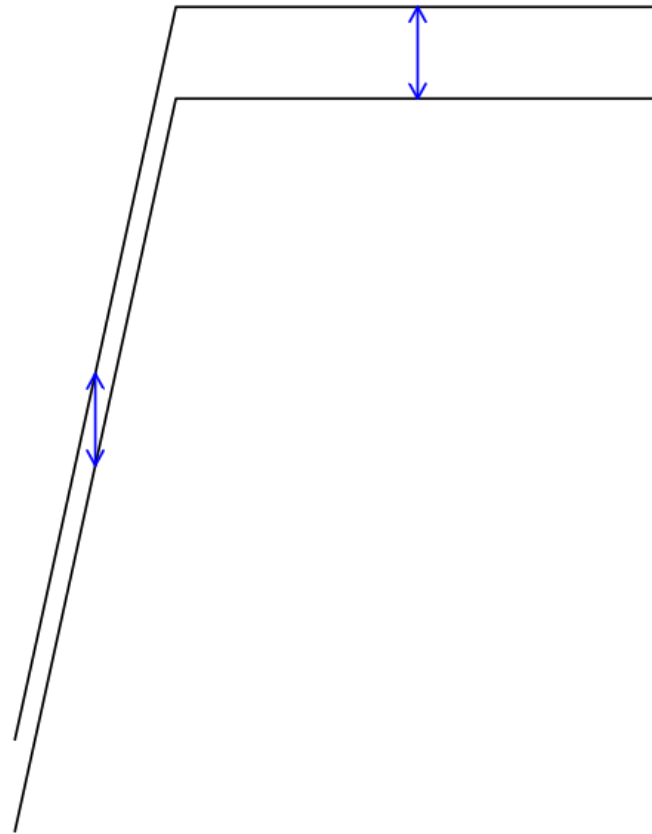


- Variables are lined up next to another
- Categories within a variable are spread out along a vertical axis
- Categories of adjacent variables are connected by boxes
- Width of boxes is proportional to the number of observations

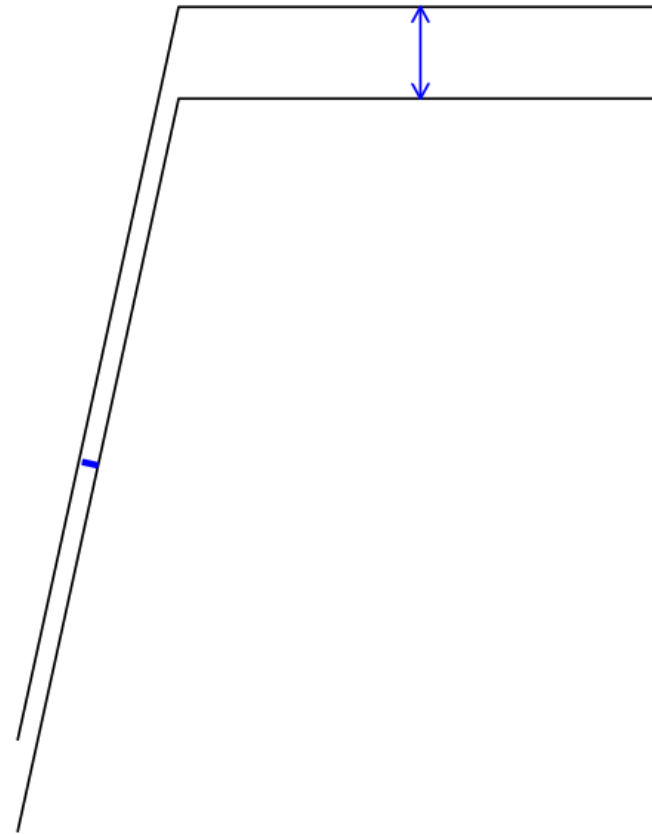
# The line-width illusion

Distance is  
perceived at a  
right angle

Your intention



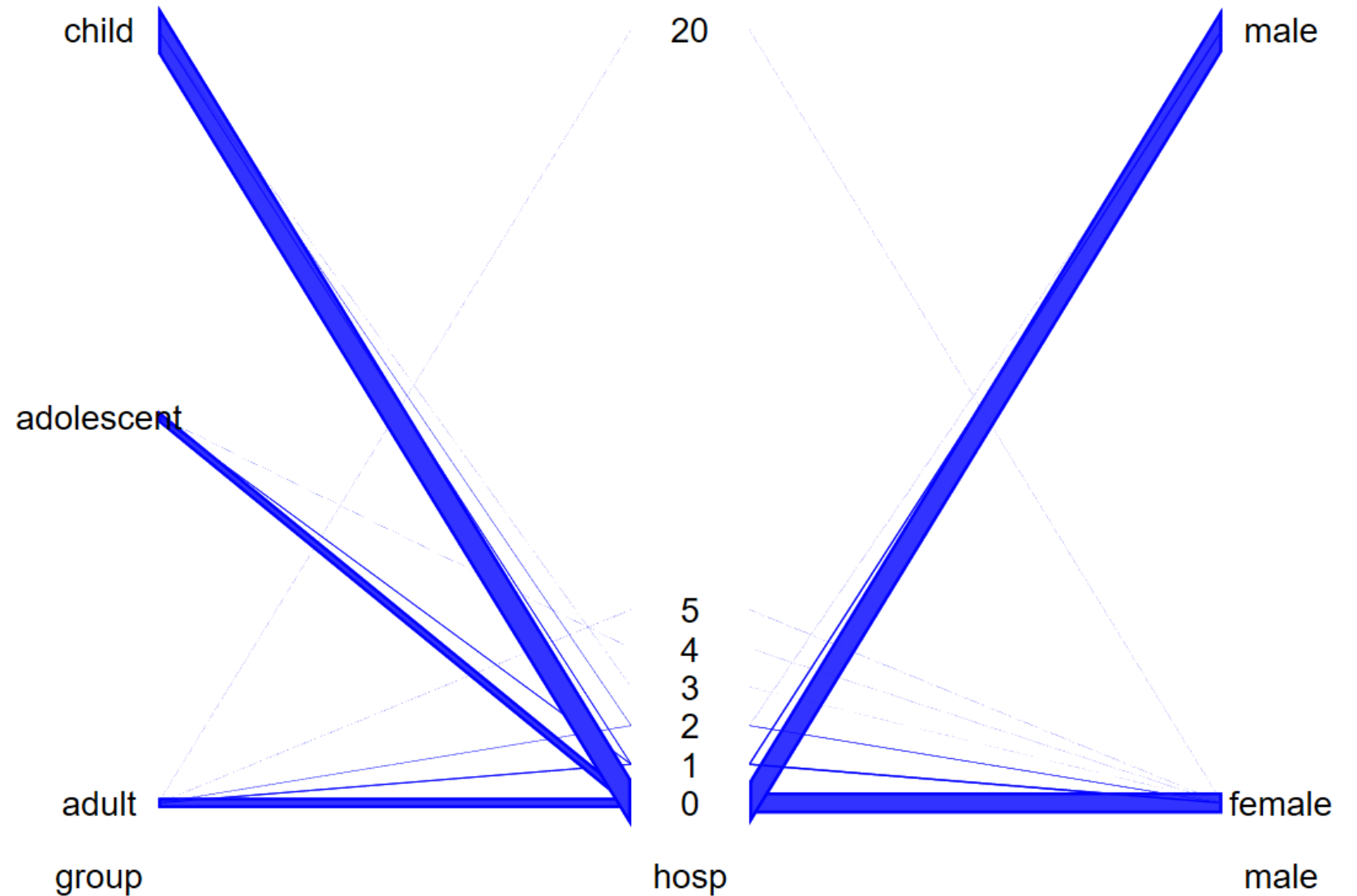
Readers' perception



e.g., Wallgren A, Wallgren B, Persson R, Jorner U, Haaland J,  
*Graphing Statistics & Data: Creating Better Charts*, Sage  
Publications, Thousand Oaks, CA. 1996.

# Hammock plot with 3 variables

- Asthma Data
- Males have never more than 2 hospital visits

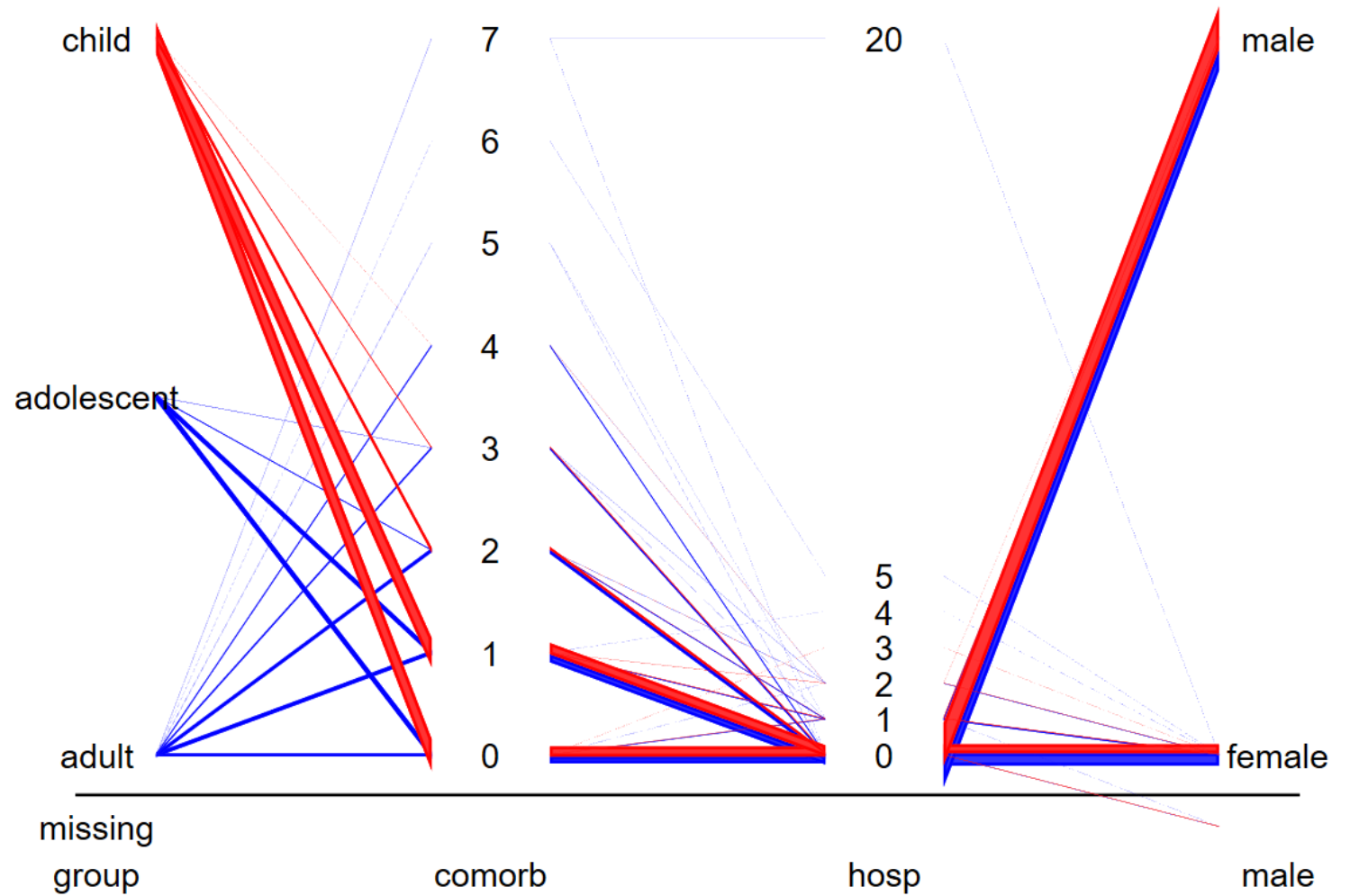


# Hammock with 4 variables

- Asthma data
- Highlighting

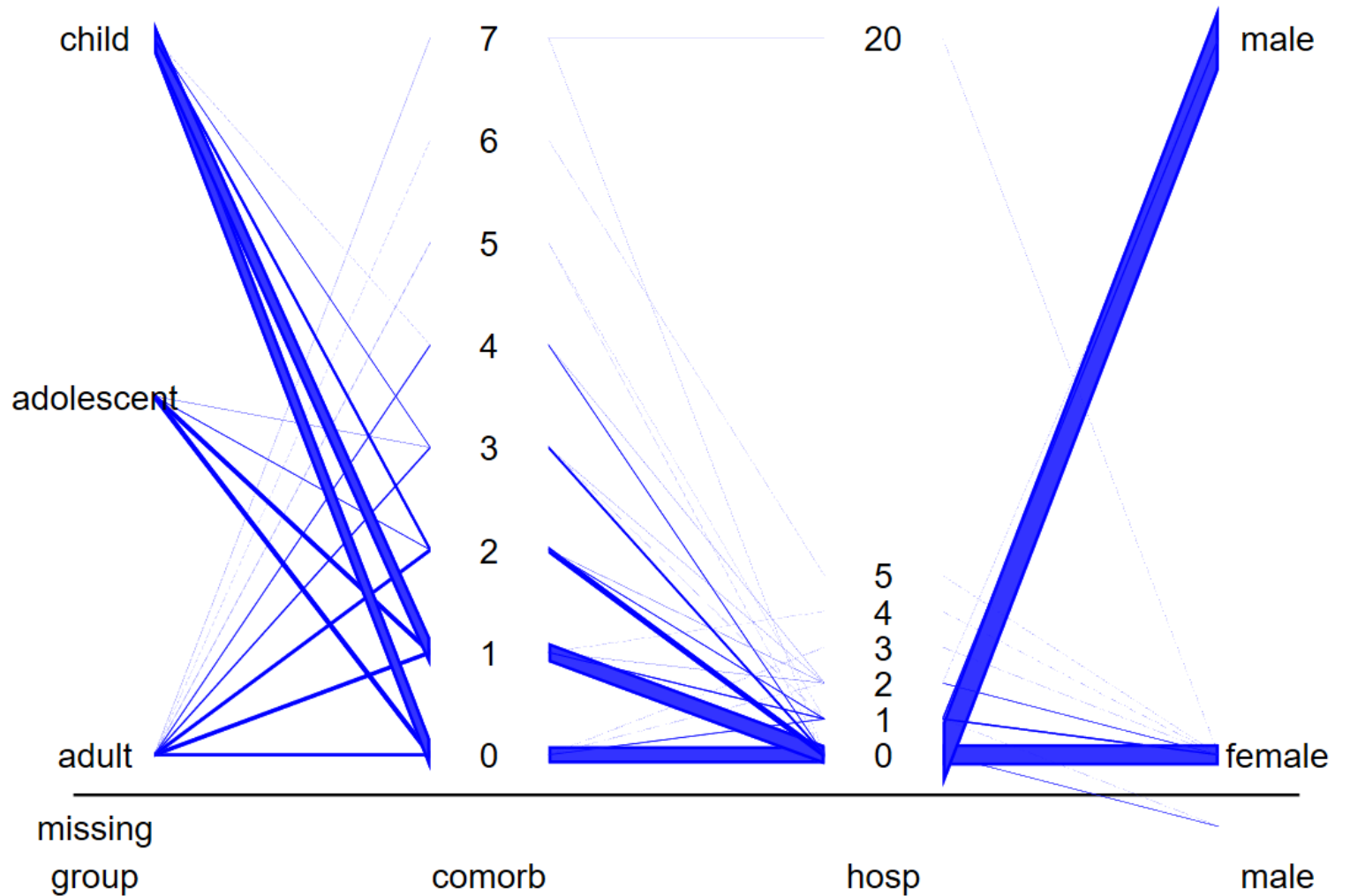
Findings:

- Adults have more comorbidities
- Can clearly identify values for outlier hosp=20



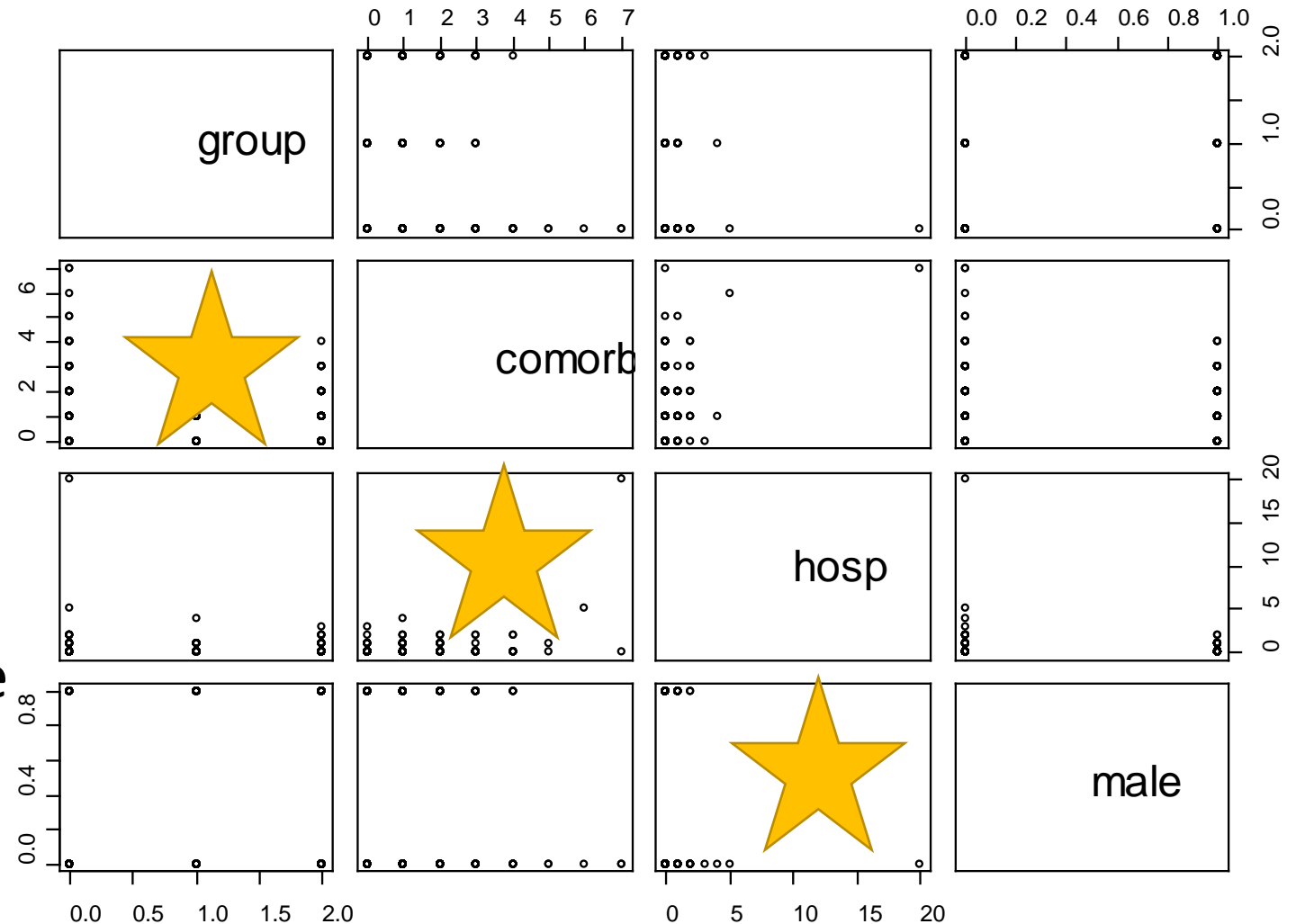
# Hammock plot with missing values

- An additional category below the line is added to track missing values
- Only the gender variable (male) has missing values



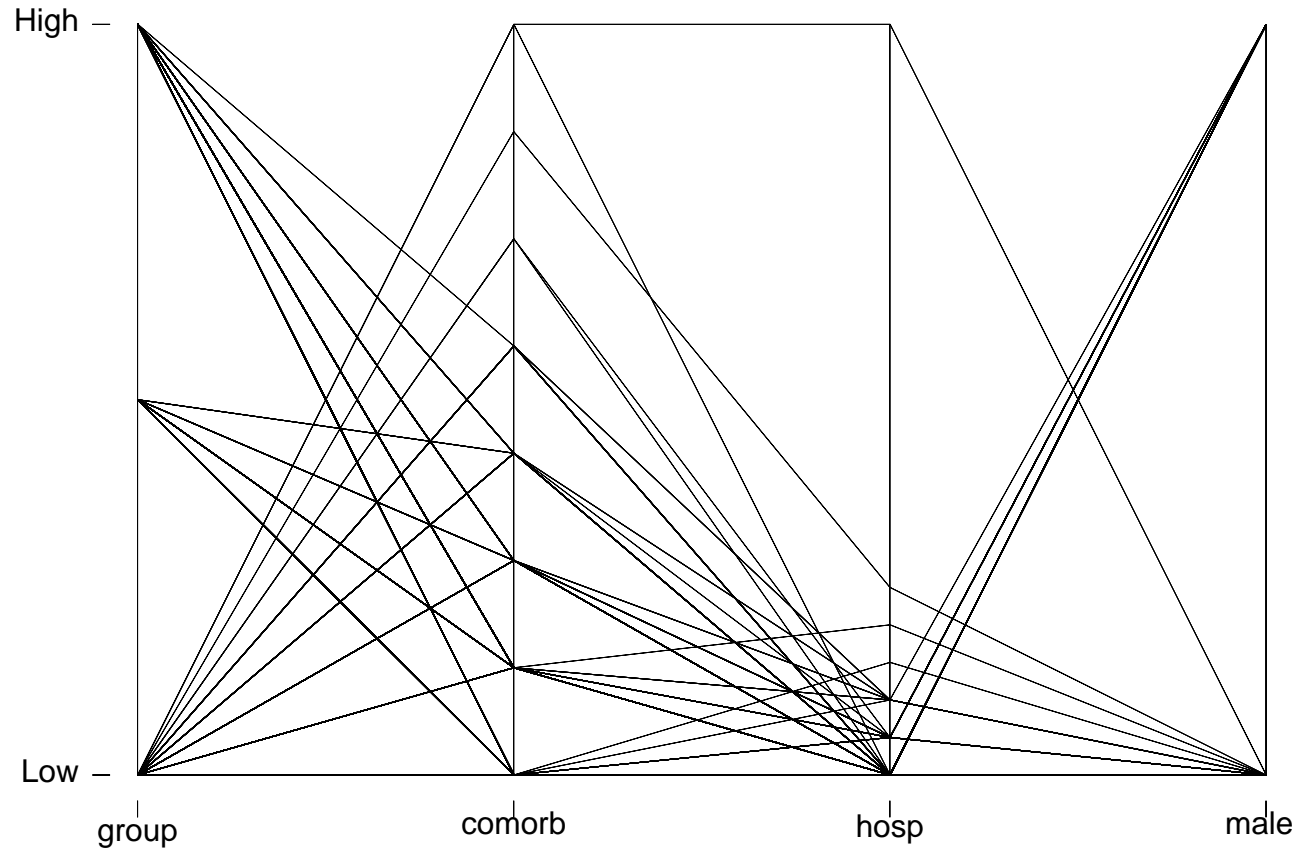
# Scatter plot matrix

- The bivariate relationships shown in hammocks and parallel coordinate plots correspond to the scatter plots adjacent to the diagonal.
- By reordering the variables, we could get the other 3



# Parallel coordinate plot

- The parallel coordinate plot emerges if you:
  - Collapse boxes into lines
  - Remove univariate descriptor
  - Remove missing value category
- For categorical variables, there is a lot of overplotting



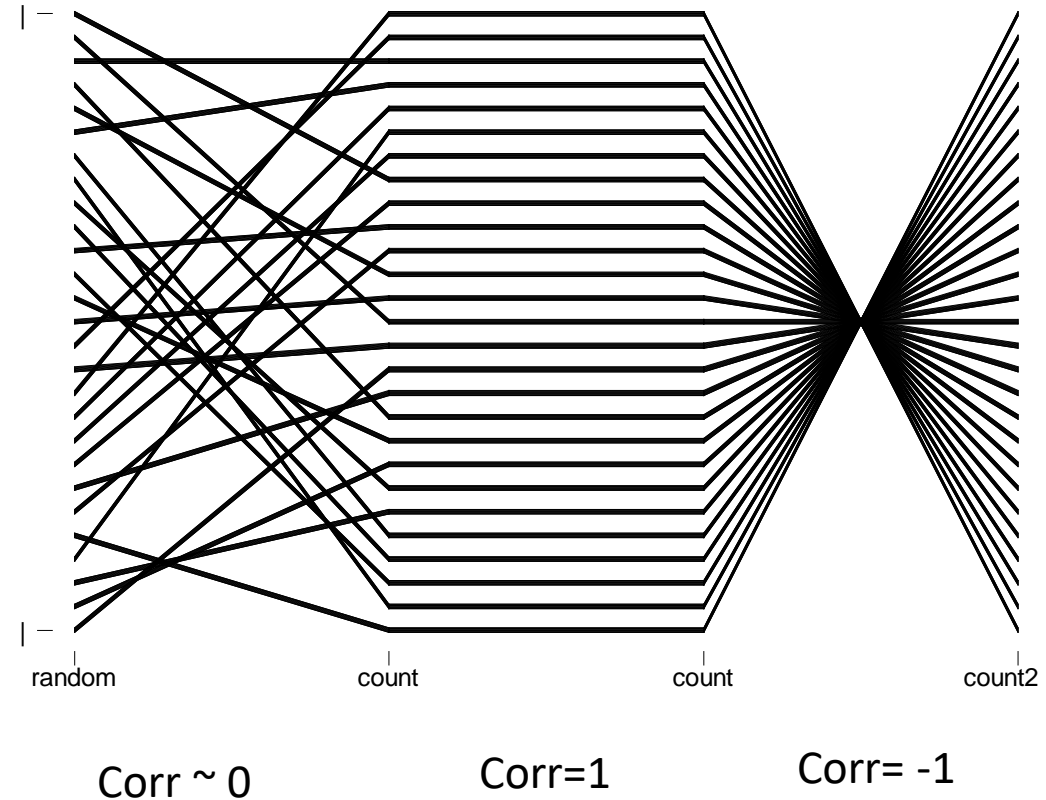
This plot was produced with the stata command:  
`parcoord group comorb hosp male`



# Correlation

If the ordering is meaningful, correlation looks as shown on the right.

If you were specifically looking for correlation, you would probably compute it directly rather than using the hammock plot.

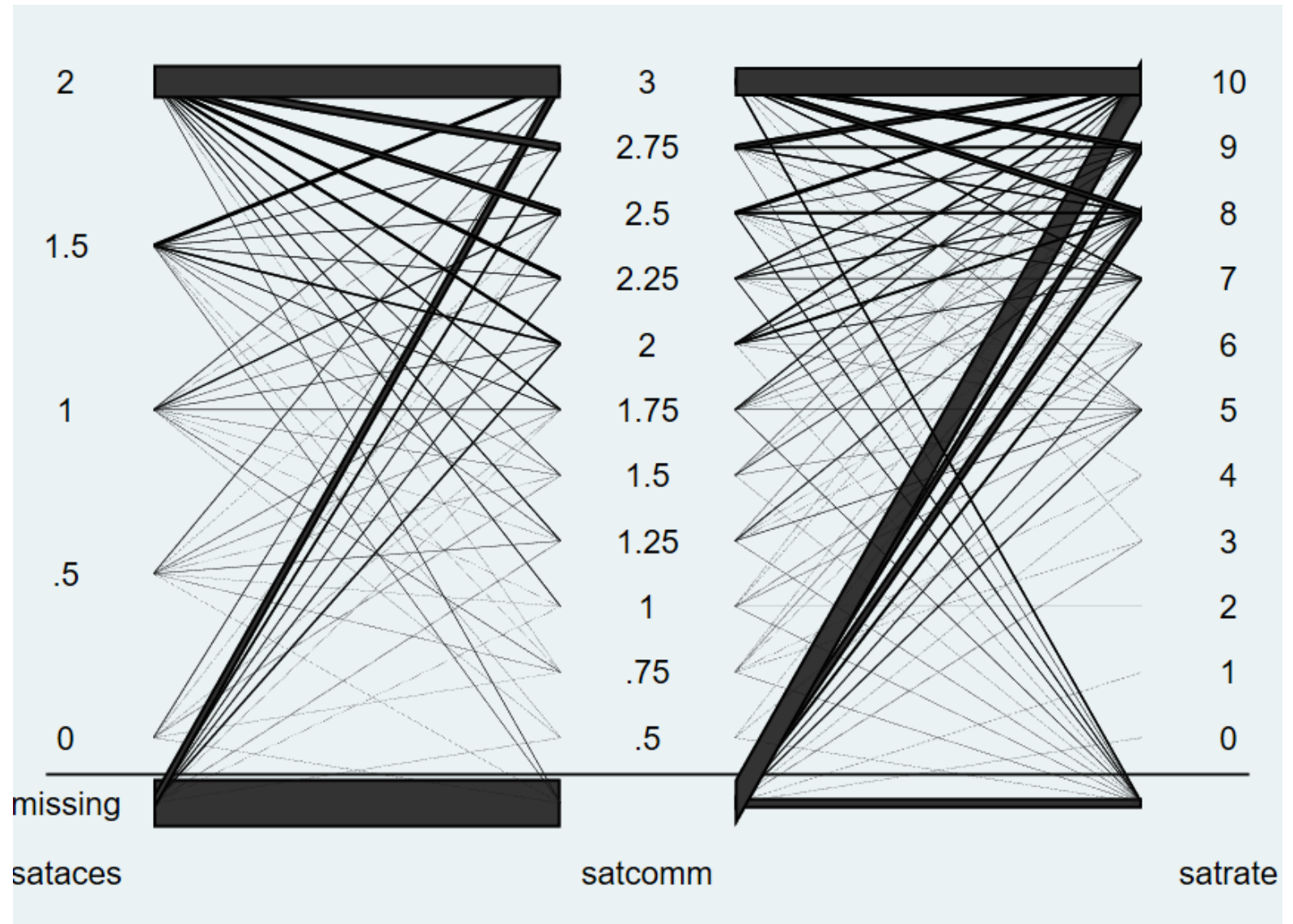


# Examples

- 1-2 slide examples
- No continuous stories of data analyses

# Example: Satisfaction Ratings in surveys

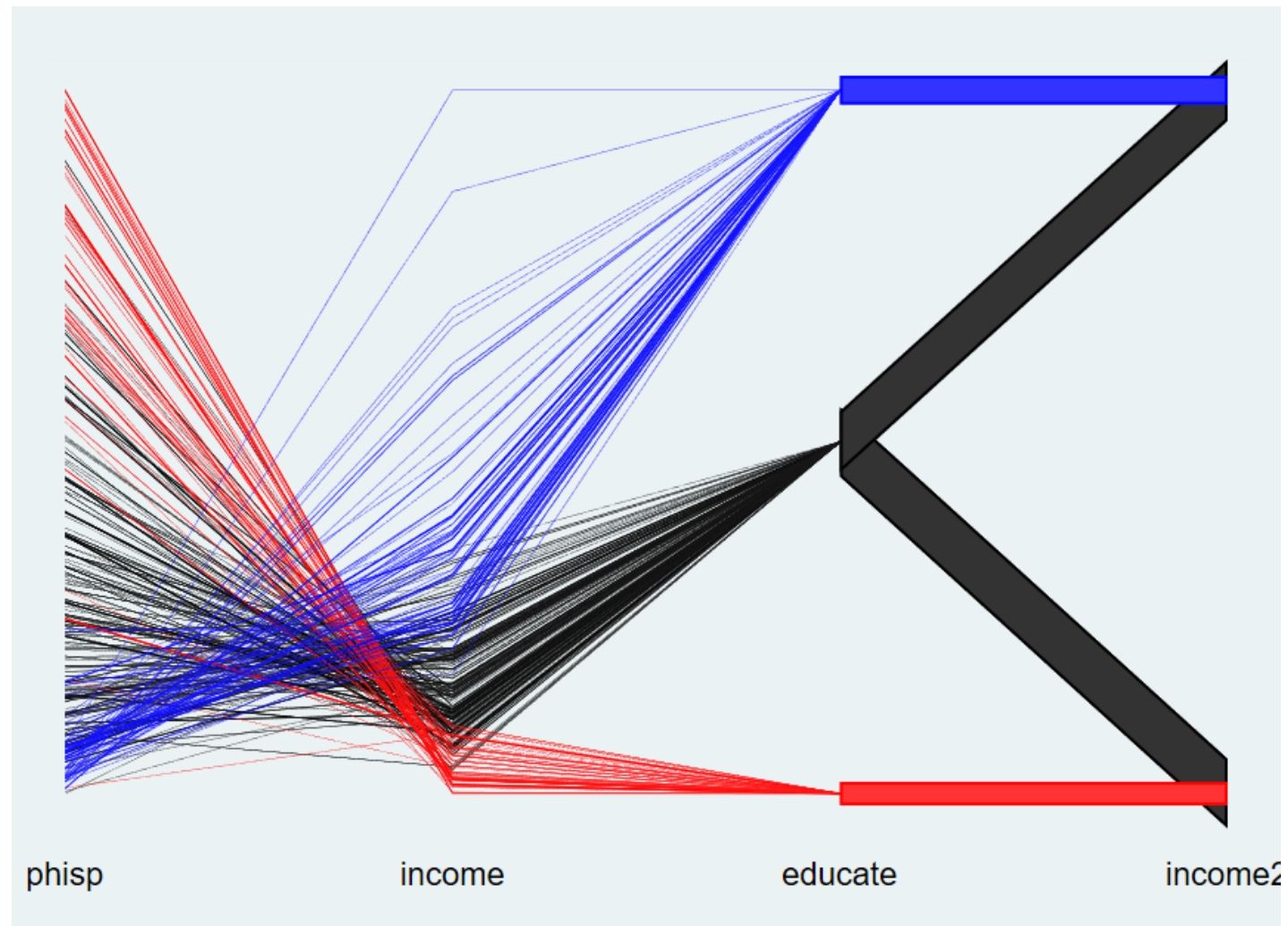
- 3 Variables:  
Satisfaction with  
access, satisfaction  
with communication,  
overall satisfaction  
rate in patients with  
diabetes
- Survey respondents  
tend to use the top  
rating



hammock sataces satcomm satrate, missing label

# Correlation

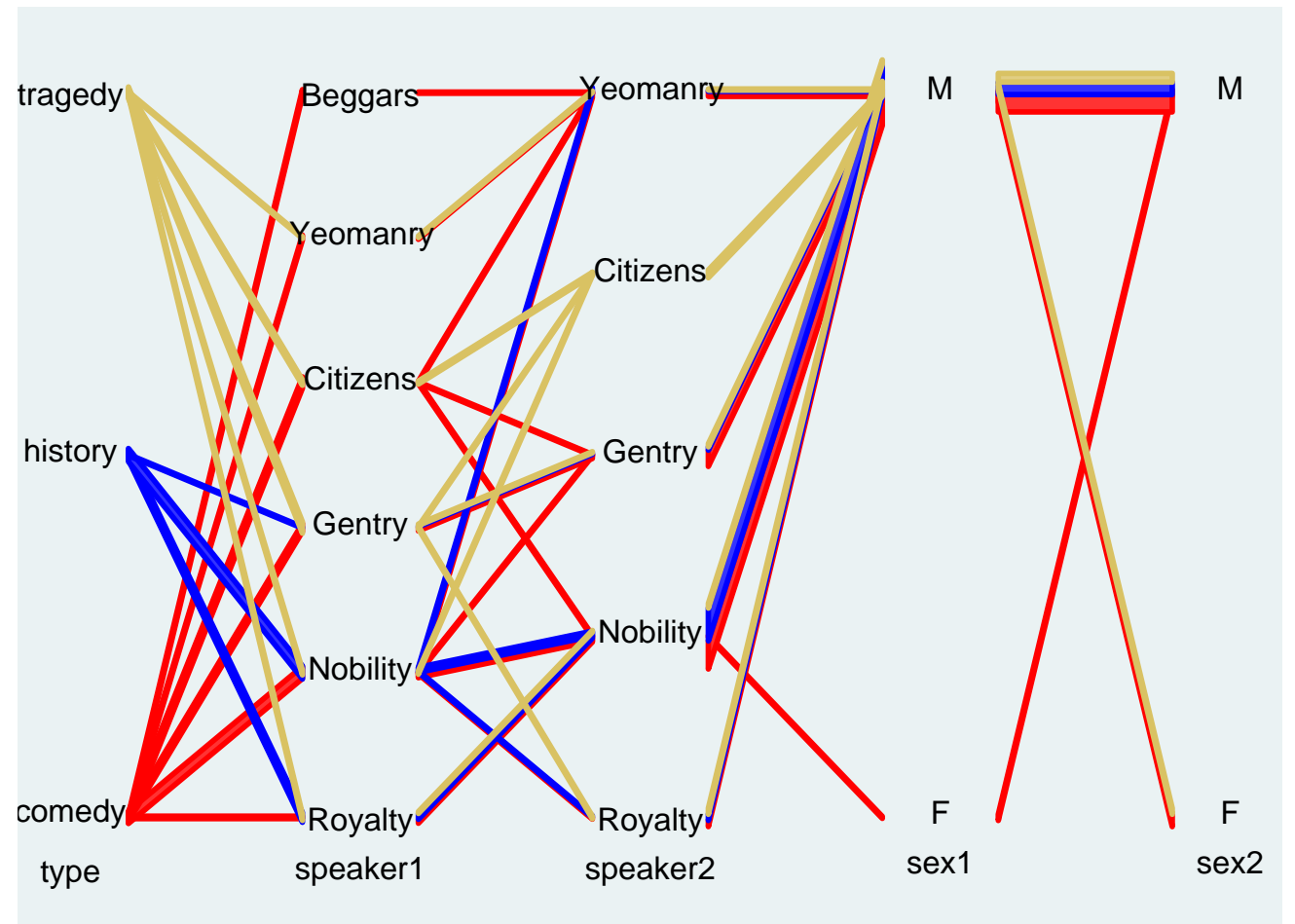
- Every observation represents a pharmacy in California
- Phisp: % Hispanics in neighborhood
- Income: median income in neighborhood
- Educate: neighborhood has low/medium/high
- Income2: hi/low depending on the ranking of income



- % Hispanics (phisp), and income are negatively correlated
- Income and education are positively correlated.

# Shakespeare data

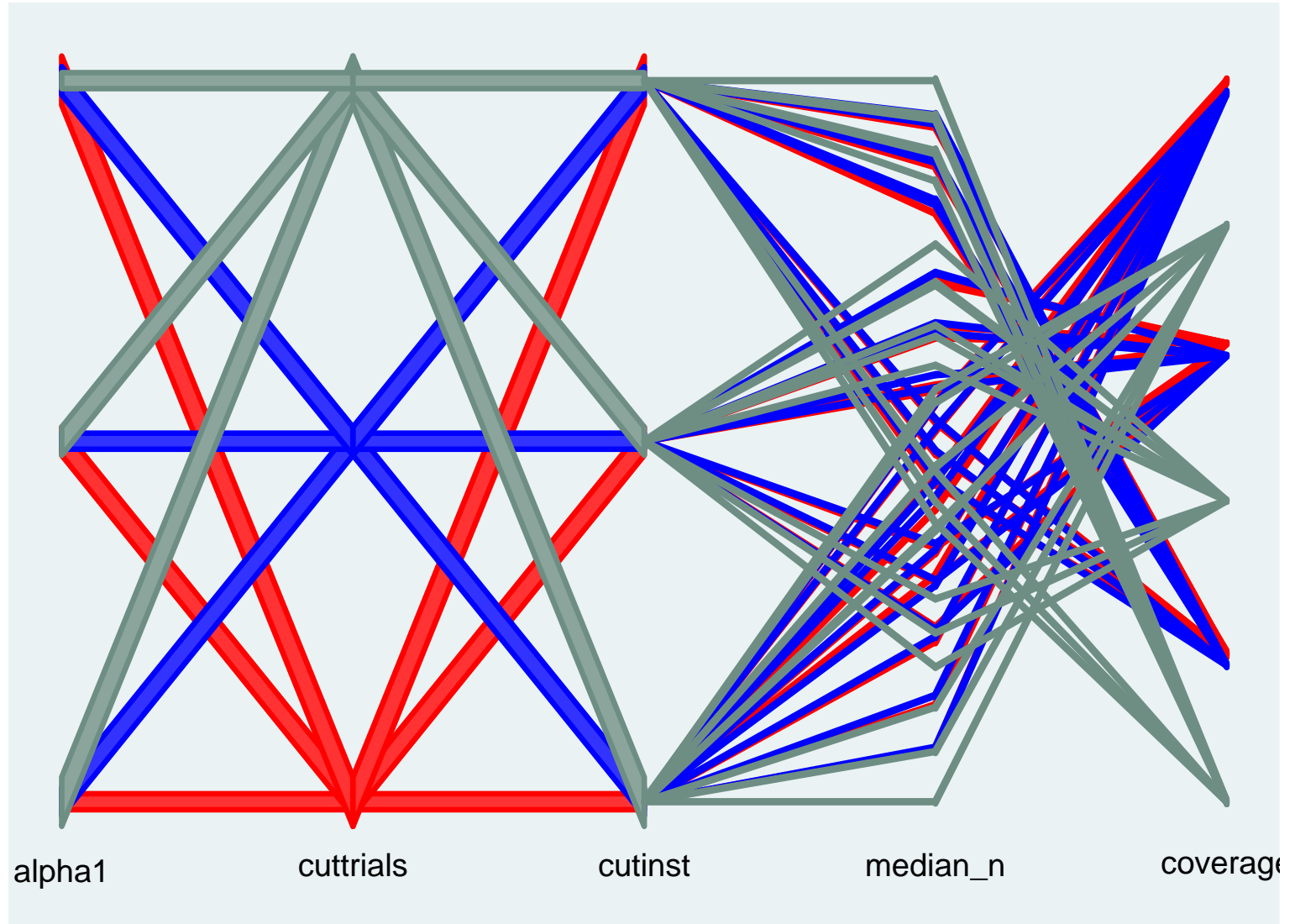
- Each observation represents a Shakespeare play
- Speaker1 and speaker2 are the 1st and 2<sup>nd</sup> persons to speak in the book
- Sex1 and sex2 are the gender of the first 2 speakers



- Most speakers are men
- Females only appear in comedies, and no female speaks to another female
- There is a roughly equal number of histories, tragedies and comedies.
- Royalty and nobility don't speak much below their station

# Example: Experimental design

- Cuttrial, cutinst, alpha1 are factors of a full factorial experiment
- cutinst (input) highly correlated with “median\_n” (output)
- Coverage highly negatively correlated with “median\_n”
- (From Cost of cancer treatment study, Dana Goldman (PI))

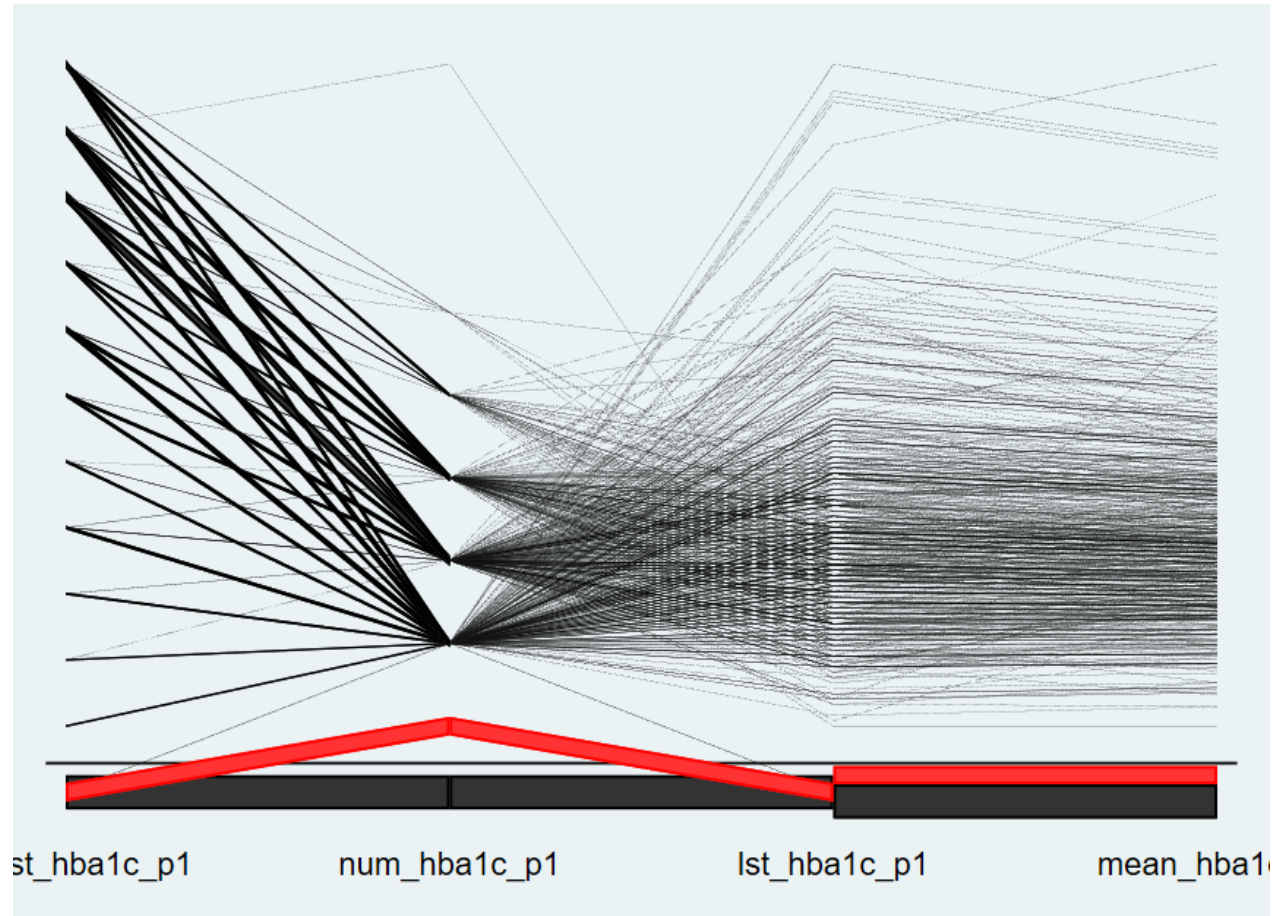


# Finding Errors

- A quick visual check helps to identify coding and other problems

# Where do missing values occur?

- The data have a lot of missing values.
- It turns out, all observations of `num_hba1c_p1` have missing values for one category





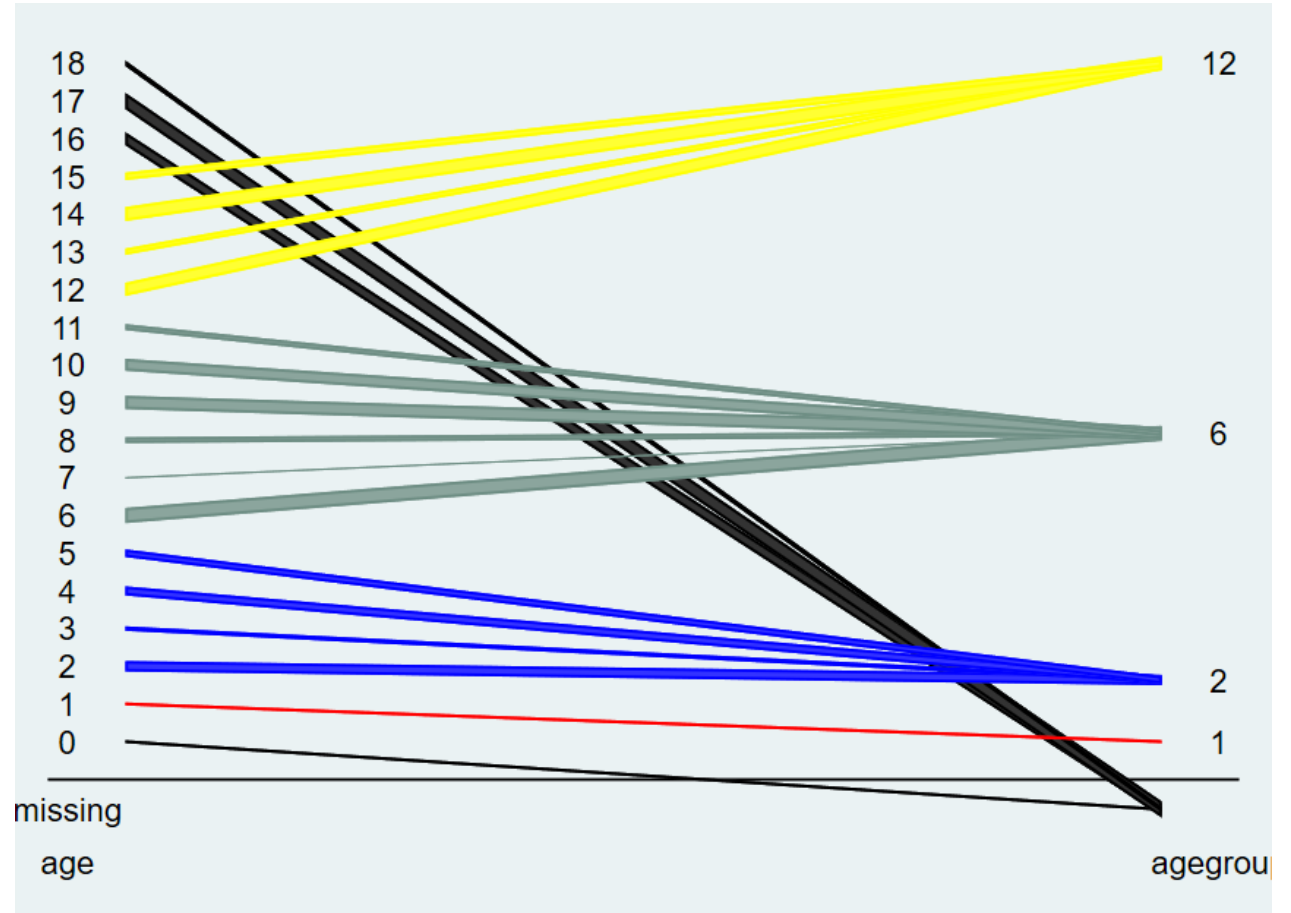
# Visual check of code

- Each observation is a person aged 0 to 18.
- Split age variable into several groups at the cut points given below:

```
egen agegroup= cut(age), at(1,2,6,12,16)
```

# Visual check of code

The visual check was a surprise:  
The oldest and youngest children's age group was missing.

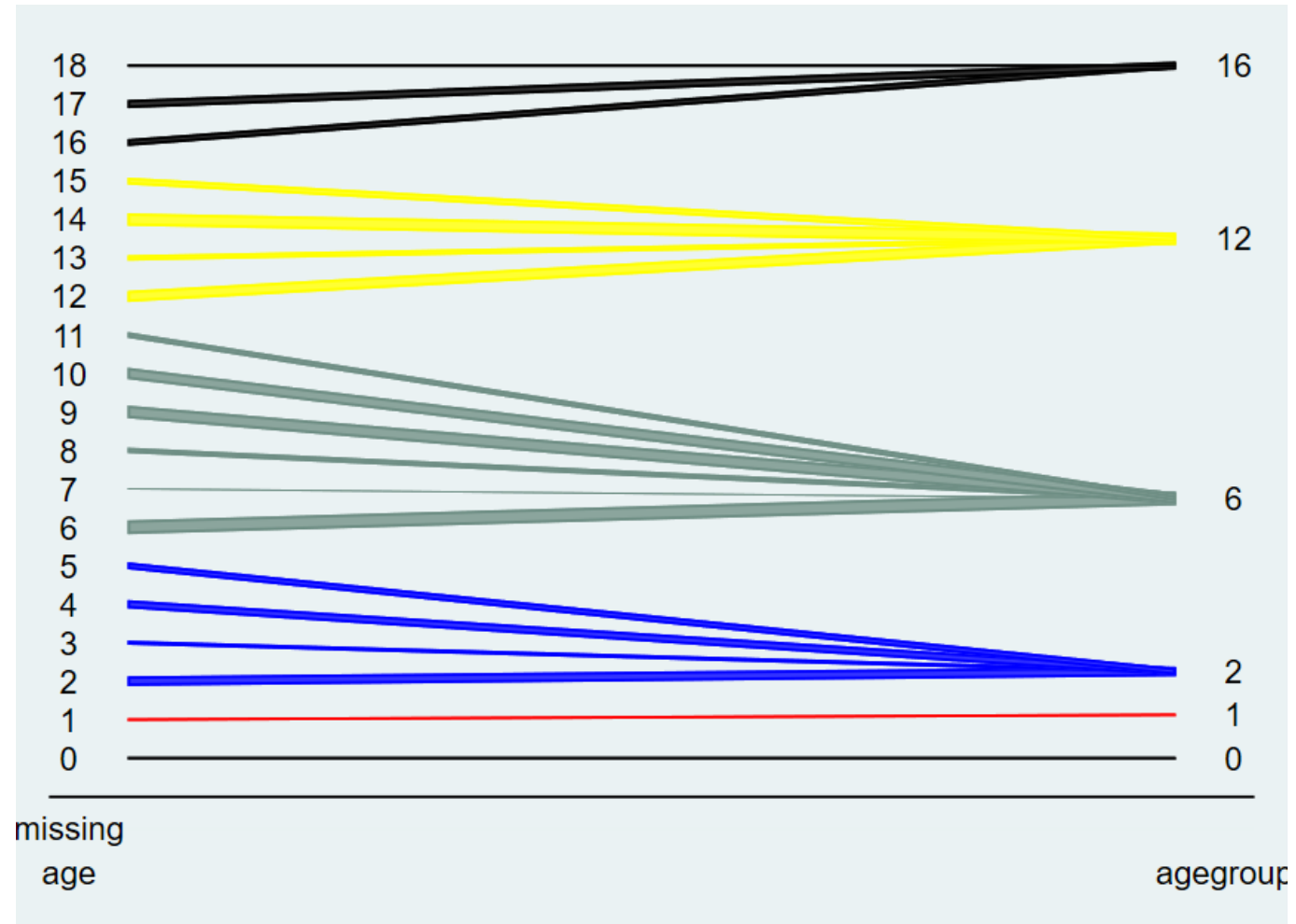


```
hammock age agegroup ,m space(.1) label hivar(agegroup) ///  
hival(1 2 6 12)
```

# Visual check of code

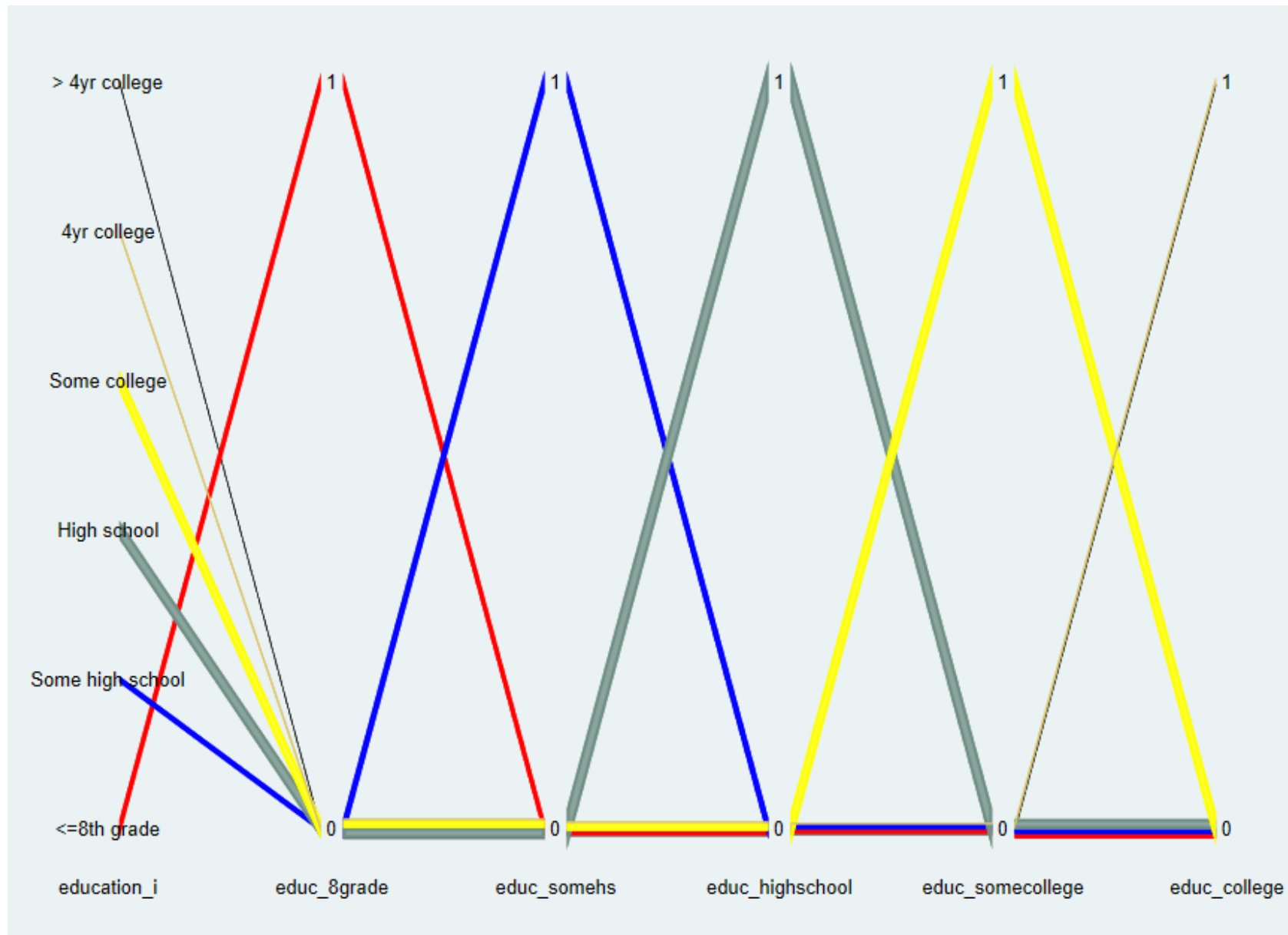
- It turned out, the command expects the end points also, and otherwise assigns missing:

```
egen agegroup2= cut(age), ///  
at(0,1,2,6,12,16,19)
```



# Example: Dummy variables

- Education variable is broken up into dummy variables
- We see that >4yr and 4yr were combined into one dummy variable



```
hammock education_i educ_8grade educ_somehs educ_highschool educ_somecollege educ_college ,  
hivar(education_i) hivalues(1 2 3 4 5) label bar(.2) spa(.1) graphregion(margin(l+5 r+1)) xlab(,labsize(vsmall))  
labelopt(size(vsmall))
```

# Exploring Multiple Imputations

- For imputation, Stata helpfiles include this example:

```
// help mi_impute
clear
webuse mheart4
tabulate alcohol, missing

//Declare data and register alcohol as imputed
mi set mlong
mi register imputed alcohol

//Impute ordered categorical variable alcohol
mi impute ologit alcohol attack smokes age bmi female hsgrad, add(5)
```

# Imputation

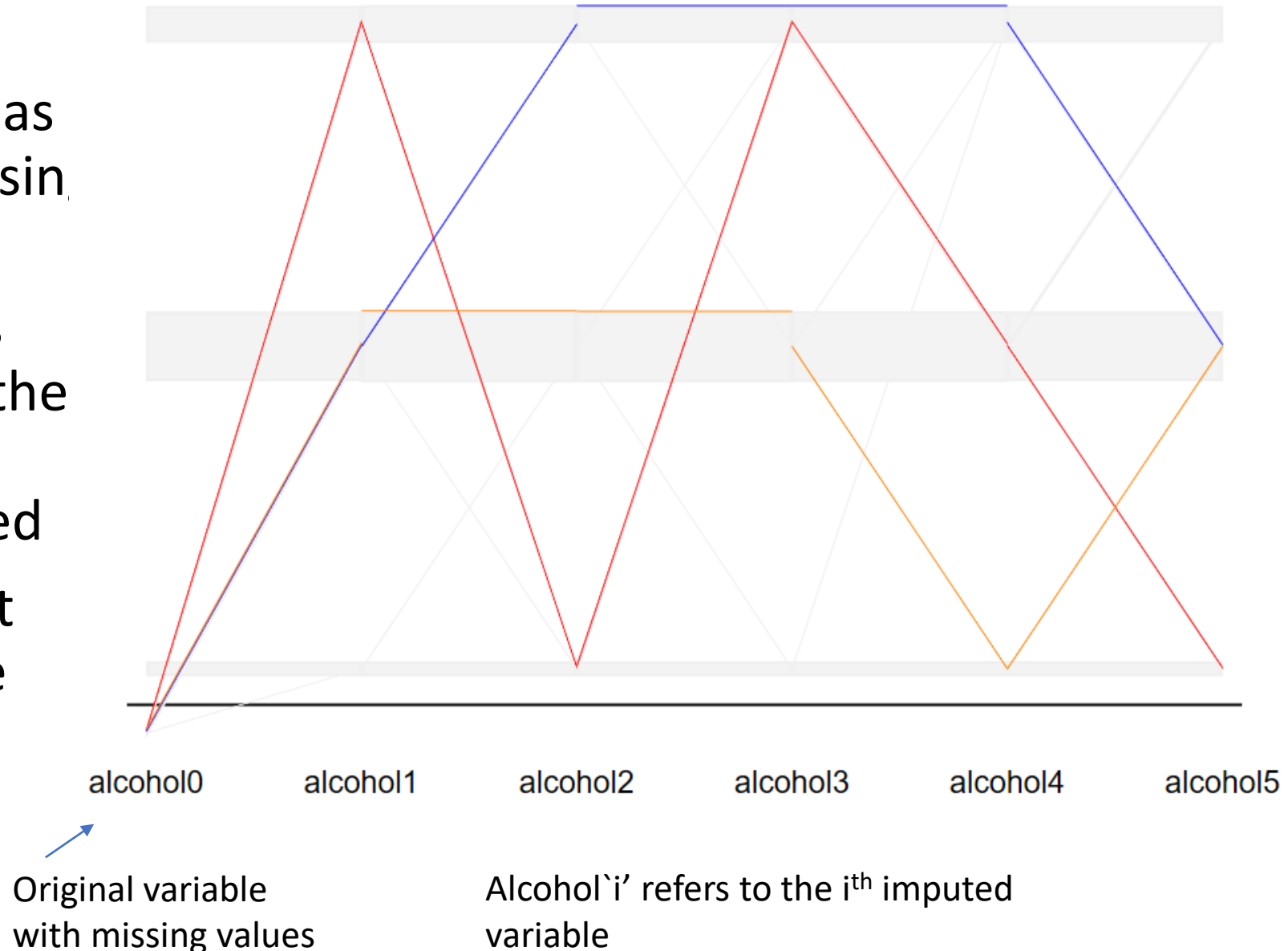
```
. tabulate alcohol, missing
```

Alcohol consumption: none, <2 drinks/day, >=2 drinks/day	Freq.	Percent	Cum.
Do not drink	18	11.69	11.69
Less than 3 drinks/day	83	53.90	65.58
Three or more drinks/day	44	28.57	94.16
.	9	5.84	100.00
Total	154	100.00	

- Do the multiple imputations actually impute different values each time?

# Imputation

- The horizontal grey areas correspond to non-missing observations.
- 3 of the 9 observations with missing values in the not-imputed variable, alcohol0, are highlighted
- The imputations do not always assign the same category



# Imputation

- Code for previous graph

```
mi unset
keep alcohol mi_m mi_id
reshape wide alcohol , i(mi_id) j(mi_m)
list mi_id if alcohol1!=.
foreach i of numlist 1/5 {
    replace alcohol`i'= alcohol0 if alcohol`i'==.
}
set scheme s1mono
hammock alcohol0 alcohol1 alcohol2 alcohol3 alcohol4 alcohol5, ///
    hivar(mi_id) hival(11 23 25 ) m col(gs15 red blue orange)///
graphregion(margin(l+3 r+1))
```

Note:

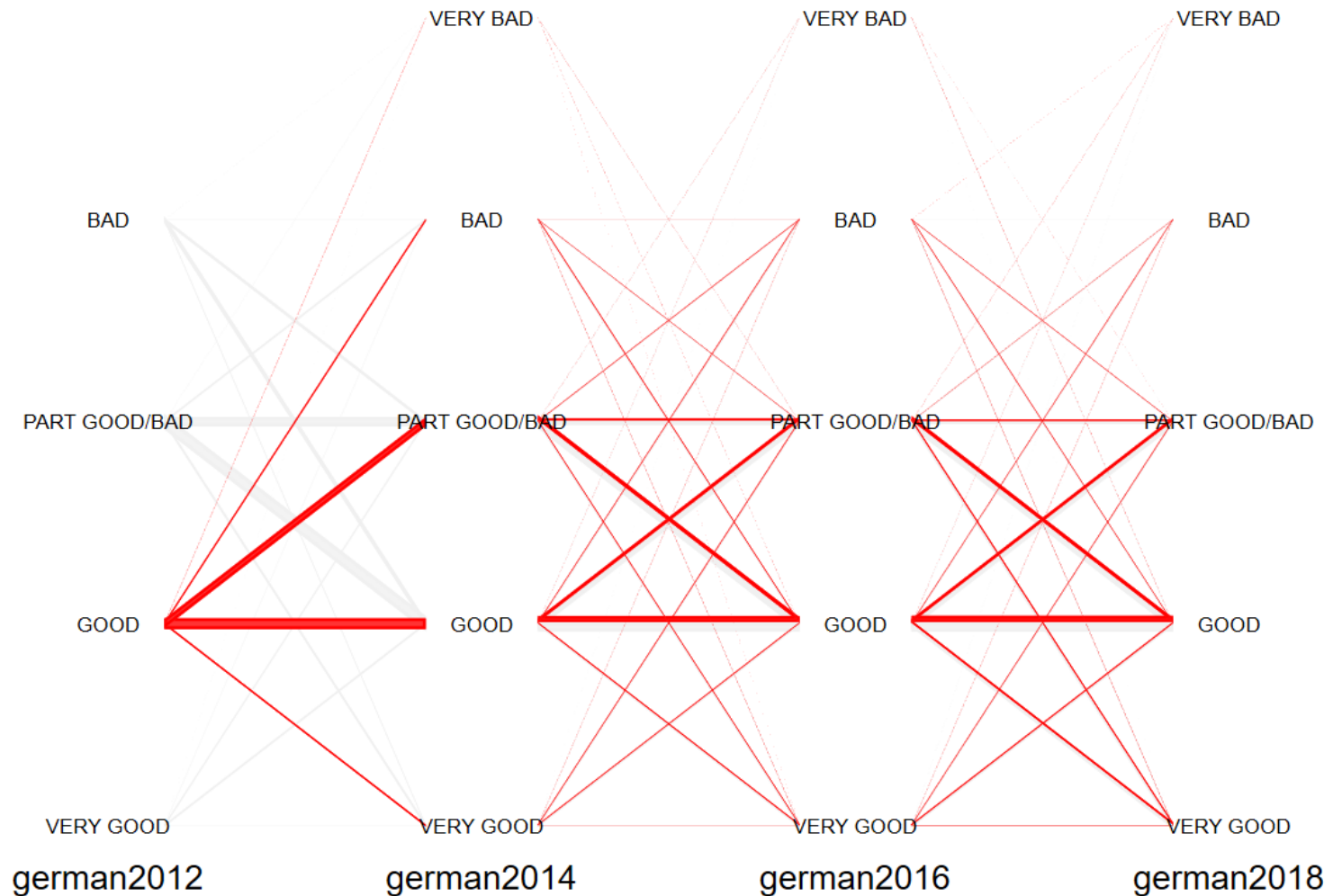
- 1) It is possible to use a highlighting variable other than the list of displayed variables.
- 2) “graphregion” expands the space at the left and right margin to display the labels better.



# What is the state of the economy in Germany?

(Source: Allbus Surveys)

- Respondent rating “good” in 2012, are all over the place in subsequence years
- The “samescale” option prevents that each variable is standardized separately  
Look at german2012



# Samescale

- Code for the previous plot

```
hammock german2012 german2014 german2016 german2018 , label ///  
  labelopt(size(vsmall)) samescale(_all) graphregion(margin(l+6 r+5)) ///  
  hivar(german2012) hival(2) col(gs15 red)
```

## Title

`hammock` — Hammock plot for visualizing categorical and continuous data

## Syntax

```
hammock varlist [if] [in] , [ options ]
```

<i>options</i>	Description
Main	
<code><u>missing</u></code>	Show missing values
<code><u>label</u></code>	Show value labels or else show values
Highlighting	
<code><u>hivariable</u>(<i>varname</i>)</code>	name of variable to highlight
<code><u>hivalues</u>(<i>numlist</i>)</code>	list of values of <i>hivariable</i> to highlight
<code><u>colorlist</u>(<i>str</i>)</code>	list of colors for highlighting
Manipulating Spacing and Layout	
<code><u>barwidth</u>(<i>real</i>)</code>	Change width of the plot elements to reduce clutter
<code><u>space</u>(<i>real</i>)</code>	Control fraction of space allocated to labels rather than to graph elements
<code><u>labelopt</u>(<i>str</i>)</code>	Pass options to <i>added_text_options</i> , e.g. to manipulate label text size
Other options	
<code><u>same</u>scale(<i>varlist</i>)</code>	Use the same axis scale for each variable specified
<code>graph_options</code>	Specify additional options passed to <i>graph</i> , <i>twoway</i>

# Related plots

- ParSets: A hammock plot where the vertical width instead of the shortest distance-width is used
- Common Angle plots: A hammock plot with an “angled” box
  - (Figure from Hofmann et al 2013)

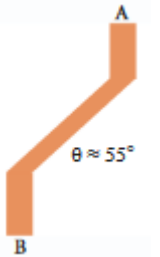
Hammock



ParSets



Common Angle plots



- Schonlau, M. (2003). Visualizing categorical data arising in the health sciences using hammock plots. In *Proceedings of the Section on Statistical Graphics, American Statistical Association*.
- Kosara, F. Bendix, and H. Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):558–568, July 2006.
- Hofmann, H., & Vendettuoli, M. (2013). Common angle plots as perception-true visualizations of categorical associations. *IEEE transactions on visualization and computer graphics*, 19(12), 2297-2305.

# Concluding Thoughts

- A graph for visualizing categorical data, but also handles continuous data
- Often not the only way of getting the information
- Graphs are vehicles of human comprehension. Everybody likes different cars.
- “In [having] the choice lies your talent.”  
(Stella Adler about acting)