Simpson’s Paradox

Jeffrey Strickland

2/22/2022

## Example

Our goal here is to look at a sitation where we can misinterpret data if our EDA is not thorough. We’ll see a phenomenon called Yule-Simpson effect or Simpson’s Paradox.

Definition. Simpson’s Paradox is a phenomenon in probability and statistics, in which a trend appears in several different groups of data but disappears or reverses when these groups are combined.

In other words, the same data set can appear to show opposite trends depending on how it’s grouped.The reason there are different interpretations of the same data, and what is evading our eye is something called the Lurking variable — a conditional variable that can affect our conclusions about the relationship between two variables.

To examine this phenomenon, we’ll use a drop-in replacement for the famous iris dataset. The dataset consists of details about three species of penguins, including their culmen length and depth, their flipper length, body mass, and sex. The culmen is essentially the upper ridge of a penguin’s beak, while their wings are called flippers. We’ll look at culmen\_length\_mm versus culmen\_depth\_mm.

## MICE Package

There are few missing values in the dataset. Let’s get rid of those. MICE (Multivariate Imputation via Chained Equations) is one of the commonly used package by R users. Creating multiple imputations as compared to a single imputation (such as mean) takes care of uncertainty in missing values.

MICE assumes that the missing data are Missing at Random (MAR), which means that the probability that a value is missing depends only on observed value and can be predicted using them. It imputes data on a variable by variable basis by specifying an imputation model per variable.

### Installe and Load Libraries

library(mice)  
library(missForest)  
library(VIM)  
library(ggplot2)  
library(readr)

### Set working directory is needed and map the path

path = getwd()  
cat("The working directory is ", path)

## The working directory is C:/Users/jeff/Documents/R/eda1

### Get zip Data from GitHub

There are two .csv files in penguin.zip \* penguin-size.csv \* penguin\_lter.csv

filename = "penguins\_size.csv"  
if (!file.exists(filename)) {  
 urlzip <- "https://github.com/stricje1/eda1/raw/main/penguins.zip"  
 download.file(urlzip, destfile = "./penguins.zip", method = "curl", extra = "-L")  
 unzip ("./penguins.zip", exdir = path )  
}

### Create dataframe from penguin-size data

df1 <- read.csv("penguins\_size.csv")  
head(df1, 11)

## species island culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm  
## 1 Adelie Torgersen 39.1 18.7 181  
## 2 Adelie Torgersen 39.5 17.4 186  
## 3 Adelie Torgersen 40.3 18.0 195  
## 4 Adelie Torgersen NA NA NA  
## 5 Adelie Torgersen 36.7 19.3 193  
## 6 Adelie Torgersen 39.3 20.6 190  
## 7 Adelie Torgersen 38.9 17.8 181  
## 8 Adelie Torgersen 39.2 19.6 195  
## 9 Adelie Torgersen 34.1 18.1 193  
## 10 Adelie Torgersen 42.0 20.2 190  
## 11 Adelie Torgersen 37.8 17.1 186  
## body\_mass\_g sex  
## 1 3750 MALE  
## 2 3800 FEMALE  
## 3 3250 FEMALE  
## 4 NA <NA>  
## 5 3450 FEMALE  
## 6 3650 MALE  
## 7 3625 FEMALE  
## 8 4675 MALE  
## 9 3475 <NA>  
## 10 4250 <NA>  
## 11 3300 <NA>

### Check for Missing Values

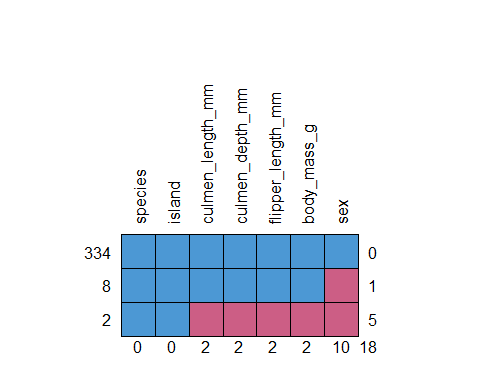
head(is.na(df1[,1:5]),8)

## species island culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm  
## [1,] FALSE FALSE FALSE FALSE FALSE  
## [2,] FALSE FALSE FALSE FALSE FALSE  
## [3,] FALSE FALSE FALSE FALSE FALSE  
## [4,] FALSE FALSE TRUE TRUE TRUE  
## [5,] FALSE FALSE FALSE FALSE FALSE  
## [6,] FALSE FALSE FALSE FALSE FALSE  
## [7,] FALSE FALSE FALSE FALSE FALSE  
## [8,] FALSE FALSE FALSE FALSE FALSE

We already knw there are missing values in the dataset, so let’s look at a visual for insight.

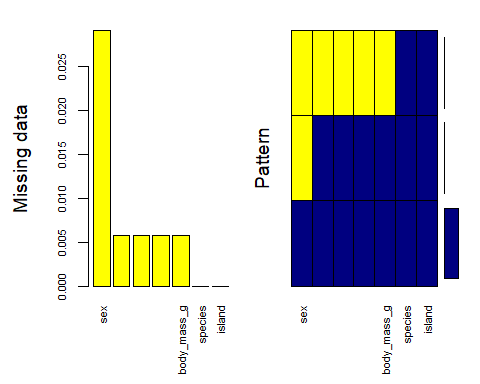
### Plot missig data patterns

md.pattern(df1, plot = TRUE, rotate.names = TRUE)



## species island culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm  
## 334 1 1 1 1 1  
## 8 1 1 1 1 1  
## 2 1 1 0 0 0  
## 0 0 2 2 2  
## body\_mass\_g sex   
## 334 1 1 0  
## 8 1 0 1  
## 2 0 0 5  
## 2 10 18

mice\_plot <- aggr(df1, col=c('navyblue','yellow'),  
 numbers=TRUE, sortVars=TRUE,  
 labels=names(df1), cex.axis=.7,  
 gap=3, ylab=c("Missing data","Pattern"))



##   
## Variables sorted by number of missings:   
## Variable Count  
## sex 0.029069767  
## culmen\_length\_mm 0.005813953  
## culmen\_depth\_mm 0.005813953  
## flipper\_length\_mm 0.005813953  
## body\_mass\_g 0.005813953  
## species 0.000000000  
## island 0.000000000

### Impute missing values

We now use the mice package to input missing values. The mice() function can impute mixes of continuous, binary, unordered categorical and ordered categorical data. In addition, mice can impute continuous two-level data, and maintain consistency between imputations by means of passive imputation. Among the parameters are the data, df1 in this case, m or the number of multiple imputations, with the defualt being 5. maxit is a scalar giving the number of iterations, with the default being 5. method can be either a single string, or a vector of strings with length ncol(data), specifying the univariate imputation method to be used for each column in data. we’ll use pmm, predictive mean matching (numeric data). Finally, we’ll set the random number seed to 500.

imputed\_Data <- mice(df1, m=5, maxit = 50, method = 'pmm', seed = 500)

##   
## iter imp variable  
## 1 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 1 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 1 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 1 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 1 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 2 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 2 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 2 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 2 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 2 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 3 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 3 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 3 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 3 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 3 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 4 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 4 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 4 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 4 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 4 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 5 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 5 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 5 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 5 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 5 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 6 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 6 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 6 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 6 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 6 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 7 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 7 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 7 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 7 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 7 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 8 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 8 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 8 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 8 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 8 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 9 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 9 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 9 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 9 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 9 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 10 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 10 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 10 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 10 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 10 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 11 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 11 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 11 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 11 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 11 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 12 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 12 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 12 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 12 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 12 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 13 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 13 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 13 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 13 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 13 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 14 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 14 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 14 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 14 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 14 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 15 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 15 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 15 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 15 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 15 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 16 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 16 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 16 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 16 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 16 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 17 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 17 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 17 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 17 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 17 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 18 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 18 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 18 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 18 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 18 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 19 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 19 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 19 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 19 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 19 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 20 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 20 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 20 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 20 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 20 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 21 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 21 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 21 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 21 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 21 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 22 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 22 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 22 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 22 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 22 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 23 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 23 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 23 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 23 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 23 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 24 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 24 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 24 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 24 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 24 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 25 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 25 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 25 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 25 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 25 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 26 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 26 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 26 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 26 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 26 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 27 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 27 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 27 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 27 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 27 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 28 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 28 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 28 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 28 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 28 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 29 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 29 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 29 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 29 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 29 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 30 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 30 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 30 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 30 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 30 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 31 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 31 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 31 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 31 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 31 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 32 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 32 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 32 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 32 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 32 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 33 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 33 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 33 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 33 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 33 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 34 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 34 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 34 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 34 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 34 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 35 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 35 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 35 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 35 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 35 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 36 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 36 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 36 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 36 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 36 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 37 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 37 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 37 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 37 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 37 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 38 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 38 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 38 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 38 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 38 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 39 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 39 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 39 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 39 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 39 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 40 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 40 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 40 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 40 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 40 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 41 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 41 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 41 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 41 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 41 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 42 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 42 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 42 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 42 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 42 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 43 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 43 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 43 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 43 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 43 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 44 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 44 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 44 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 44 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 44 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 45 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 45 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 45 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 45 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 45 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 46 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 46 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 46 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 46 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 46 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 47 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 47 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 47 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 47 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 47 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 48 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 48 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 48 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 48 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 48 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 49 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 49 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 49 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 49 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 49 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 50 1 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 50 2 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 50 3 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 50 4 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g  
## 50 5 culmen\_length\_mm culmen\_depth\_mm flipper\_length\_mm body\_mass\_g

summary(imputed\_Data)

## Class: mids  
## Number of multiple imputations: 5   
## Imputation methods:  
## species island culmen\_length\_mm culmen\_depth\_mm   
## "" "" "pmm" "pmm"   
## flipper\_length\_mm body\_mass\_g sex   
## "pmm" "pmm" ""   
## PredictorMatrix:  
## species island culmen\_length\_mm culmen\_depth\_mm  
## species 0 0 1 1  
## island 0 0 1 1  
## culmen\_length\_mm 0 0 0 1  
## culmen\_depth\_mm 0 0 1 0  
## flipper\_length\_mm 0 0 1 1  
## body\_mass\_g 0 0 1 1  
## flipper\_length\_mm body\_mass\_g sex  
## species 1 1 0  
## island 1 1 0  
## culmen\_length\_mm 1 1 0  
## culmen\_depth\_mm 1 1 0  
## flipper\_length\_mm 0 1 0  
## body\_mass\_g 1 0 0  
## Number of logged events: 3   
## it im dep meth out  
## 1 0 0 constant species  
## 2 0 0 constant island  
## 3 0 0 constant sex

### Check Imputed Dataset

Now that we have imputed missing values, we want to verify that we were successful. The next code chunk shows us the values that replaced the missing rows (4 and 430) for each of the 5 imputations.

imputed\_Data$imp$culmen\_length\_mm

## 1 2 3 4 5  
## 4 46.6 38.8 41.6 39.7 42.5  
## 340 49.5 50.5 39.7 35.0 37.3

From the output, we’ll select the third imputation values to replace the missing values.

### Get complete data (3rd out of 7)

The complete() function takes an object of class mids, fills in the missing data, and returns the completed data in a specified format. To repeat, we’ll fill in the third imputation (recall that we did m = 5 imputations.

completeData <- complete(imputed\_Data,3)

### Recheck for missing values

Now we’ll reinspect the data for missing values. The output shows there are none.

is.na(completeData)==0

### Create dataframe

Now, we’ll create a datframe from the complteData imputed data.

df2 <- data.frame(completeData)

### Define x and y variables

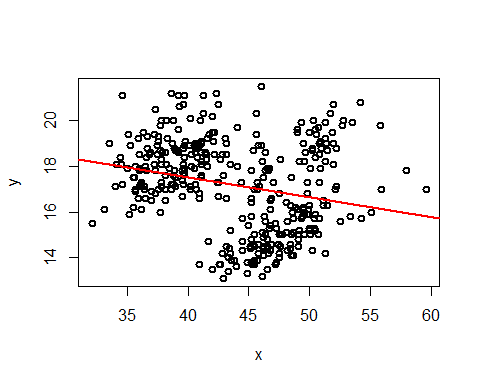
Here, we’ll define the varoables x and y as culmen\_length\_mm and culmen\_depth\_mm, respectively.

x <- df2[,3]  
y <- df2[,4]

### Create scatterplot

Now, we’ll generate a scatterplot of the x and y variables, with a line fitted for the data. This plot will show a trend, indicated by the fitted line, of the two plotted variables. It shows that there is a slightly downward trend (negative slope) in the data when considering culmen\_length\_mm versus culmen\_depth\_mm only. This demonstrates Simpson’s Paradox.

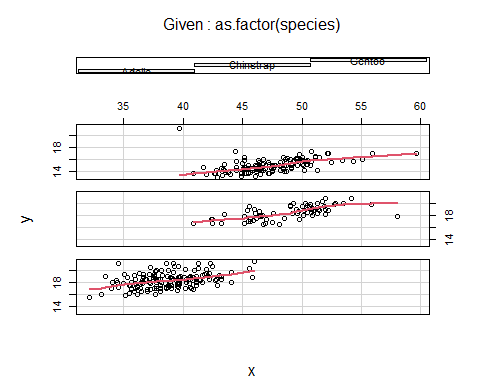
plot(x,y, lwd = 2)  
abline(lm(y ~ x, data = df2), col = "red", lwd = 2)



### Create Coplot by Species

If we stopped now, we might not understand what the data is trying to tell us as we interrogate it. Plotting culmen\_length\_mm versus culmen\_depth\_mm by the three penguin species yields a different insight. Here we can see that the actual trends are slightly positive.

require(graphics)  
coplot(y~x | as.factor(species), data = df2,  
 panel = panel.smooth, rows = 3, lwd = 2)

 ### Conclusion