Project Report G9

Jan Alexander, Paul Morbée, Joren Verbeke, Steven Wallaert

2019 - 2020

Contents

1 Introduction

2 Methods

2.1 Analysis protocol

2.2 Data cleaning and preparation

```
filename <- "armpit.txt"
armpit <- read.table(filename)
summary(armpit)
```

```
##  Corynebacterium.1 Corynebacterium.2 Corynebacterium.3 Corynebacterium.4
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.2019   Median : 0.3458   Median : 0.0000   Median : 0.0000
##  Mean   : 6.5259   Mean   :11.4825   Mean   : 1.9110   Mean   : 1.2008
##  3rd Qu.: 1.6563   3rd Qu.: 2.7544   3rd Qu.: 0.7368   3rd Qu.: 0.2539
##  Max.   :96.1716   Max.   :87.1904   Max.   :30.2198   Max.   :18.3894
##
##  Staphylococcus.1  Staphylococcus.2  Staphylococcus.3 Staphylococcus.4
##  Min.   :  1.203   Min.   : 0.0000   Min.   :0.0000   Min.   : 0.00000
##  1st Qu.: 42.155   1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.: 0.00000
##  Median : 93.504   Median : 0.0000   Median :0.0000   Median : 0.00000
##  Mean   : 71.173   Mean   : 6.2755   Mean   :0.4602   Mean   : 0.97152
##  3rd Qu.: 98.961   3rd Qu.: 0.2506   3rd Qu.:0.0801   3rd Qu.: 0.07669
##  Max.   :100.000   Max.   :89.7331   Max.   :6.6667   Max.   :21.56593
```

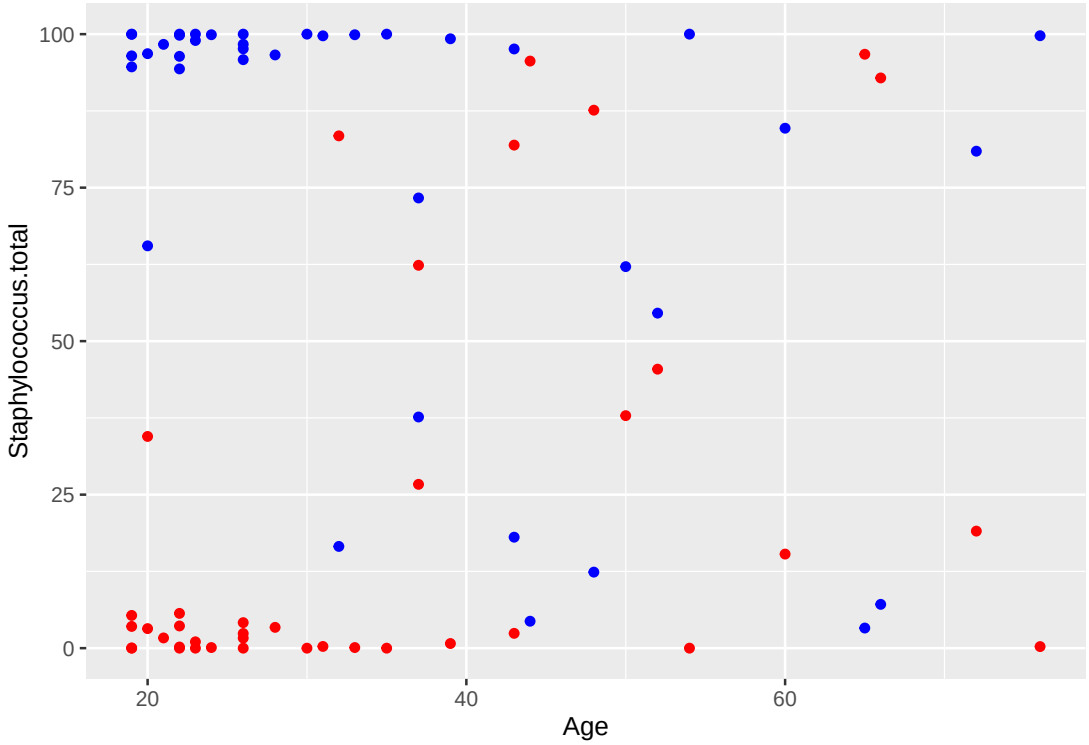
```
##
##       Age              BMI         Gender
##  Min.   :19.00   Min.   :0.000      : 1
##  1st Qu.:22.00   1st Qu.:0.000   F :23
##  Median :30.00   Median :0.000   F : 1
##  Mean   :35.23   Mean   :0.325   M :14
##  3rd Qu.:43.50   3rd Qu.:1.000   M : 1
##  Max.   :76.00   Max.   :1.000
##  NA's   :1
```

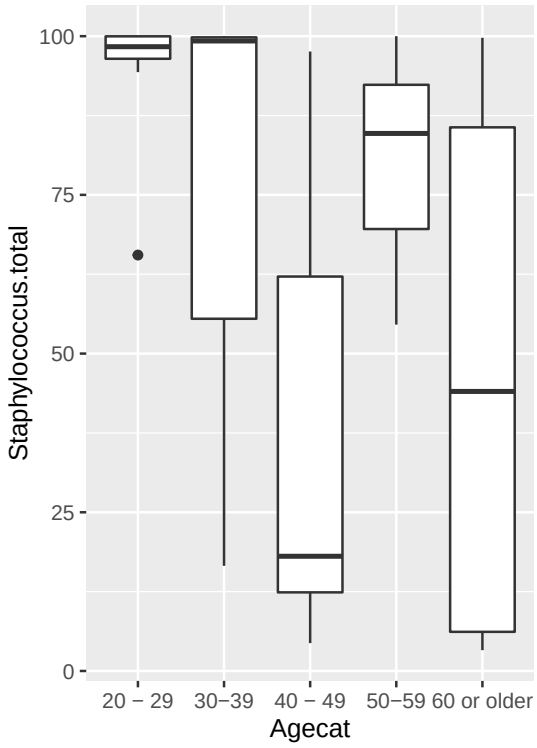
```
# Step 1: assure we have correct gender labels
# convert the gender 'F ' to 'F' and ' M' to 'M'
armpit[grepl('F', armpit$Gender), 'Gender'] <- 'F'
armpit[grepl('M', armpit$Gender), 'Gender'] <- 'M'
# drop the observation where gender and age were not filled out --> TODO : is this acceptable?
armpit <- armpit[armpit$Gender != '',]
armpit$Gender <- factor(armpit$Gender)
armpit <- armpit[!is.na(armpit$Age),]
summary(armpit)
```

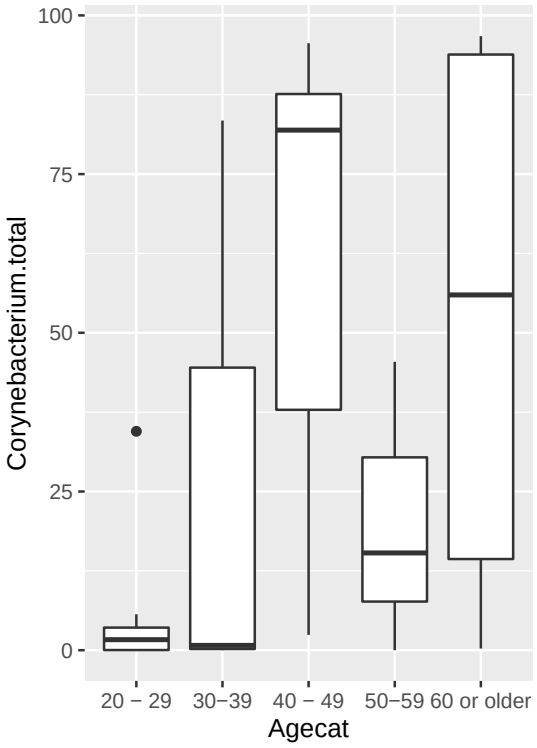
```
##  Corynebacterium.1 Corynebacterium.2 Corynebacterium.3 Corynebacterium.4
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.000    Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.000    1st Qu.: 0.0000
##  Median : 0.1968   Median : 0.2011   Median : 0.000    Median : 0.0000
##  Mean   : 6.6481   Mean   :11.7397   Mean   : 1.396    Mean   : 1.2183
##  3rd Qu.: 1.6086   3rd Qu.: 3.0679   3rd Qu.: 0.540    3rd Qu.: 0.1478
##  Max.   :96.1716   Max.   :87.1904   Max.   :30.220    Max.   :18.3894
##  Staphylococcus.1  Staphylococcus.2  Staphylococcus.3  Staphylococcus.4
##  Min.   :  1.203   Min.   : 0.0000   Min.   :0.00000   Min.   : 0.00000
##  1st Qu.: 40.355   1st Qu.: 0.0000   1st Qu.:0.00000   1st Qu.: 0.00000
##  Median : 93.675   Median : 0.0000   Median :0.00000   Median : 0.00000
##  Mean   : 71.213   Mean   : 6.3247   Mean   :0.46404   Mean   : 0.99643
##  3rd Qu.: 99.022   3rd Qu.: 0.1605   3rd Qu.:0.06568   3rd Qu.: 0.08635
##  Max.   :100.000   Max.   :89.7331   Max.   :6.66667   Max.   :21.56593
##       Age              BMI         Gender
##  Min.   :19.00   Min.   :0.0000   F:24
##  1st Qu.:22.00   1st Qu.:0.0000   M:15
##  Median :30.00   Median :0.0000
##  Mean   :35.23   Mean   :0.3333
##  3rd Qu.:43.50   3rd Qu.:1.0000
##  Max.   :76.00   Max.   :1.0000
```

3 Results

3.1 Correlation between bacteria and age







4 Conclusion