

Stat 579 - Homework #1

Your Name

9/6/2018

Swiss Banknotes

1. Download the RMarkdown file with these homework instructions to use as a template for your work. Make sure to replace “Your Name” in the YAML with your name.
2. The R package `alr3` contains a data set called `banknote`, consisting of (physical) measurements on 200 Swiss bank notes, 100 of which are genuine, while the other half is counterfeit. Load this data set (you might have to install the package) using the code below. Also run the cryptic third line - it will make your life a lot easier for the rest of the homework. This turns variable `Y` explicitly into a factor variable, i.e. makes it discrete. We will discuss this in the course material later in more detail.

```
# install.packages("alr3")
library(alr3) # if this throws an error of the form 'there is no package called alr3', uncomment the li

## Warning: package 'alr3' was built under R version 3.4.4
## Warning: package 'car' was built under R version 3.4.4
## Warning: package 'carData' was built under R version 3.4.4

data(banknote)
banknote$Y <- factor(banknote$Y)
```

3. Use one of our object inspecting functions and interpret the result in the data that you see.
4. Use the package `ggplot2` to draw a barchart of `Y` (0 is genuine, 1 is counterfeit). Map `Y` to the fill color of the barchart.
5. Use the package `ggplot2` to draw a histogram of one of the variables in the dataset that shows a distinction between genuine and counterfeit banknotes. Use fill color to show this difference. Choose the binwidth such that there are no gaps in the middle range of the histogram.
6. Use the package `ggplot2` to draw a scatterplot of two (continuous) measurements, color by `Y`. Try to find a pair of measurements that allow you to separate perfectly between genuine and counterfeit banknotes.
7. For each of the three figures above, write a two-three sentence summary, describing the
 - a. **structure of the plot:** what type of plot is it? Which variables are mapped to x, to y, and to the (fill) color?
 - b. **main message of the plot:** what is your main finding, i.e. what do you want viewers to learn from the plot?
 - c. **additional message:** point out anomalies or outliers, if there are any.

Due date: the homework is due a week from now on Thursday.

For the submission: submit your solution in an R Markdown file and (just for insurance) submit the corresponding html/word file with it.