Case 1 'CALUX'

<-- Your Name Goes Here-->

2017-05-19

## Directions

*Use ggplot syntaxis for all the figures in this case study.*

*Use this Rmarkdown template to create a report adressing the assignment below. The Rmd template can be found in the folder: //home/../cases/*

## Introduction

For the assessment of carcinogenic potential of chemicals, animal testing required. Recently, we have performed a ring trial with two laboratories to acertain the predictive capacity of a reporter cell line to accurately classify chemicals on the basis of reporter signal (fluorescence).

## The case assignment

1. Create a github repository "case\_1\_calux" in your account.
2. Read the data into R
3. Inspect the data
4. Describe mising values and the type of variables
5. Generate at least 4 different graphs that describe the data.
6. Answer the case questions below.
7. Upload a rendered (github\_document) and the Rmd file to you github account
8. Sent the teacher an email with the link to the files

## **TIPs**

* For this case you need to know about importing data in R. The file is csv format, remember to set the na.strings. Determine the delimiter used in the datafile. You can open the file in a text editor.
* Remember the programming rules: use snake\_case fore all variables in the data and in your script.
* Use the syntax from the {tidyverse} as much a possible
* You are allowed to use google/stack overflow, whatever you can find to solve the questions
* All graphics must be made with {ggplot2}
* Remember {dplyr} to select(), filter(), group\_by(), mutate() and summarize() data
* Check the type of each variable in the data. Do you need to change the variable type like data$var\_1 <- as.factor(data$var\_1)?

## Case questions

Try to answer the next 4 questions:

Graph 1) How does the relationship between the chemicals ($name) and the $p53\_cyto\_mec variable look?

Graph 2) Can you plot the relationships between the $class variable and the $p53.ifmax variable for each $laboratory?

Graph 3) Make a graph with facets that display only chemical sthat have an induction of mean p53.ifmax of above 5.0.

Graph 4) Create a meaningful bar-graph for this data set. You can choose what to plot yourself

## Review criteria

* Your code must provide the steps above in such a way that someone else can understand the code.
* Provide a clear description of what your code does
* Describe the graphs in your report.
* Provide the complete code in a rendered Rmd (to HTML) file

## Case data

Load the data from the data folder "//root/data/case\_1\_calux/"

This folder contains two files:

*variables\_calux.md* and *zonmw\_calux\_all\_data.csv*

The data for this case is in the file "zomw\_calux\_all\_data.csv"

You need to change the names of the variables to the "conventions" (e.g. $p53\_cyto\_mec) do this first after loading the data.

library(tidyverse)  
data\_path <- file.path(root, "data", "case\_1\_calux", "zonmw\_calux\_all\_data.csv")  
prevalidation <- read\_delim(data\_path,   
 na = "NA", delim = ";")

## Adding row\_id

Remember "Data management" it is good idea to add a row id

## Change names of variables

## Case question that will help you create the above 4 graphs:

## 1) Geoms

The chemicals tested were divided into 3 classes. This graph is about showing the realtionship btween the chemicals and the induction of p53. Can you think of a way to convey this information in a plot? What type or types of geom\_ will you use?

## 2) The variables in the data

Which variable types are represented in the data. You can find more information in the "variables\_calux.md" file. Change the type of the variable if necessary.

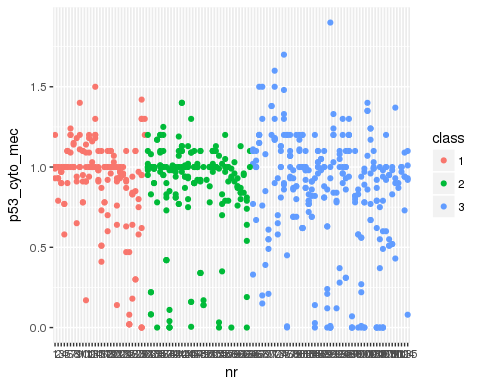
## 3) Complete cases

The dataset has a number of missing values and incomplete cases. First filter the data for all the cases that are complete

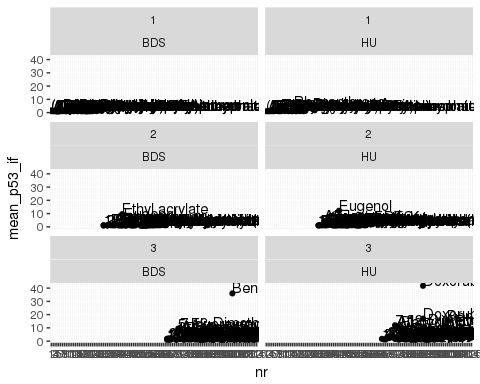
## 3) Summarize data

Do you need to calculate means for every chemical? In oder to investigate relations it is probably a good idea to calculate the mean of the duplos (you can see two observations for most chemical). Calculate a summary where you group\_by(laboratory, name, nr, class, s9\_mix) Use this summary for graph 2-4.

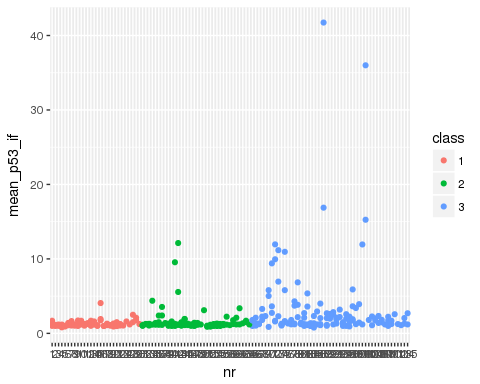
## Example plot 1:

How does the relationship between the chemicals ($name) and the $p53\_cyto\_mec variable look? 

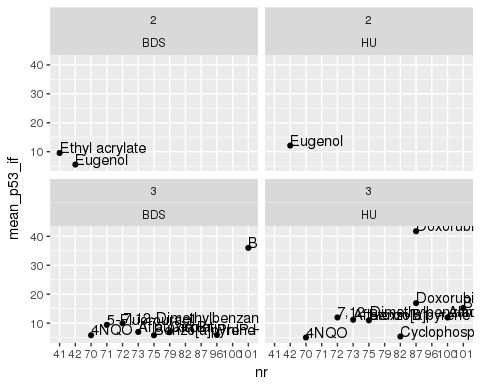
## Example plot 2:

Can you plot the relationships between the $name, the $class variable and the $p53.ifmax variable for each $laboratory? Make this plot from summarized data (sumarize the observation for each duplo measurement) 

## Another example



## Example graph 3

Make a graph with facets that display only the chemical that have an induction of mean p53\_if of above 5.0. 

What can you conclude from this plot on the relation between class of the chemical and the induction of p53\_if\_max?

Example graph 4 Create a meaningful bar-graph for this data set. You can choose what to plot yourself See e.g. page 29 from "r4ds" 