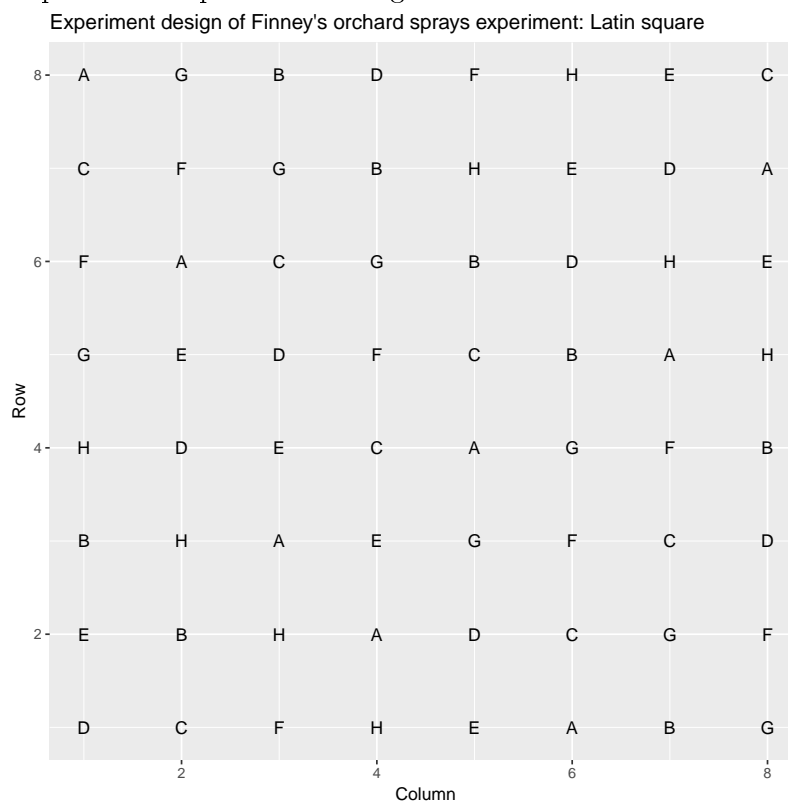


- (0.25 points) 1. Create a vector x with the integer values 5, 17, 11, 3 and 14. Write down the R command.
- (0.25 points) 2. Create a vector y with continuous integer values from 2 to 76. Write down the R command.
- (0.25 points) 3. Name four data types in R .
- (0.25 points) 4. Name three data structures in R .
- (2 points) 5. Create the dataframe "A" with the following data and commands:
- ```
> students <- c("Jerry", "Julia", "Jason", "Estelle", "Micheal")
> ages <- c(67, 60, 62, 93, 72)
> heights <- c(1.8, 1.6, 1.64, 1.63, 1.9)
> A <- data.frame(name=students, age=ages, height=heights, row.names=1)
```
- (0.5 points) (a) Add an additional column "gender" with the corresponding abbreviations to each name (f=female, m=male) and recreate the dataframe "A". Write down the  $R$  commands.
- (0.5 points) (b) Calculate the average age of women and men separately (use the **tidyverse** package if you like).
- (0.5 points) (c) Calculate the average height of all people older than 65 (use the **tidyverse** package if you like).
- (0.5 points) (d) Create a subset of the dataframe "A", only containing row one and three in one  $R$  command.
- (0.5 points) 6. Create a vector with 40 random data points from the normal distribution with a standard deviation of 75. Plot a histogram with the data and draw the mean and median as a vertical line in the graph (Use the **ggplot2** package if you like). Keep in mind: You get will get a new (i.e. different) data set every time you call the function to generate random data points.
- (1.5 points) 7. Use  $R$ 's built in dataset *OrchardSprays* (use the **ggplot2** package if you like).
- (0.25 points) (a) Briefly describe the 'OrchardSprays' experiment based on the information in the  $R$  help. Here is a plot of the design:



- (0.25 points) (b) Create a boxplot of the decrease for each treatment A-H.
- (0.25 points) (c) Add a meaningful title and axis labels
- (0.25 points) (d) Create a histogram for treatment G
- (0.5 points) (e) Create a boxplot for each row of the Latin Square design. Each row should contain data from all treatments. What is the expectation based on the design: Should there be differences between the rows or not? What do you conclude from your plot?

This exercise has 7 questions, for a total of 5 points.