

## Homework 4 - ggplot 2 package and graphical data representation

The present Homework consists of three short exercises on the use of ggplot to produce graphical data representation. The aim is to identify if you understand how plots are organized and if you can reproduce graphs using ggplot, aesthetic mapping, and layers.

### 0.1 Exercises

**Exercise 0.1** For each plot below, indicate what are the data, aesthetic mappings and layers. You will need to guess a little because you have not seen all the data sets and functions yet, but use your common sense! See if you can predict what the plot will look like.

1. `ggplot(mpg, aes(cty, hwy)) + geom_point()`
2. `ggplot(diamonds, aes(carat, price)) + geom_point()`
3. `ggplot(economics, aes(date, unemploy)) + geom_line()`
4. `ggplot(mpg, aes(cty)) + geom_histogram()`

**Exercise 0.2** Replicate the graph Figure 1. Use the in-built dataset `mtcars`.

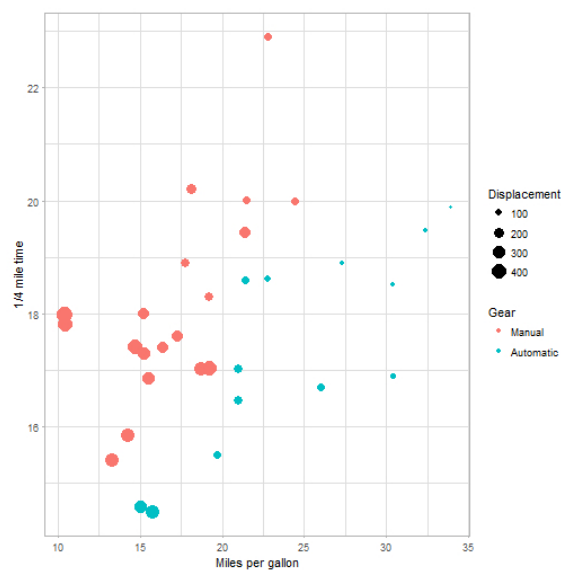


Figure 1: Graph to obtain in exercise 2

**Exercise 0.3** In this exercise we are going to visualize the changes in US births between 1978 and 2015. The data are available in the package `mozaic`.

1. Install the package `mozaic` and load it into your R script
2. Check for each data frame **Births78** and **Births2015** what are the numbers of observations and variables composing it, and also the format of each variable.
3. Combine the data frames **Births78** and **Births2015** in such a way that the number of columns in the combined data frame is the same than in the individual data frame. Assign the combined data frame to a variable named **USBirths78vs2015**.
4. What are the names of the variables in the **USBirths78vs2015**?
5. How many rows and columns does **USBirths78vs2015** have?
6. Plot a scatterplot of number of births per year day split by year (see the graph Figure 2).  
textitTip1: transform the variable **year** into a factor before performing the plot. **Tip2: to choose your own colors, look at how to use `scale_color_manual()`**
7. Create a new variable **positionInWeek** in the data frame **USBirths78vs2015**, that takes the value **Weekend** if the birth has occurred on a Sunday or a Saturday, and the value **Working day** otherwise. *Tips: you can use a **ifelse** in your recode the factor*
8. Plot a bar chart that shows for each year the number of births, split by **positionInWeek**. See the graph Figure 3 for the expected output.
9. Plot a jitter-plot plotting for each month the numbers of births, split by position in the week, and divided by year. See the graph Figure 4 for the expected output. *Tip: use the function **facet\_grid()**, and the function **theme()** to change the direction of the labels on the x-axis*



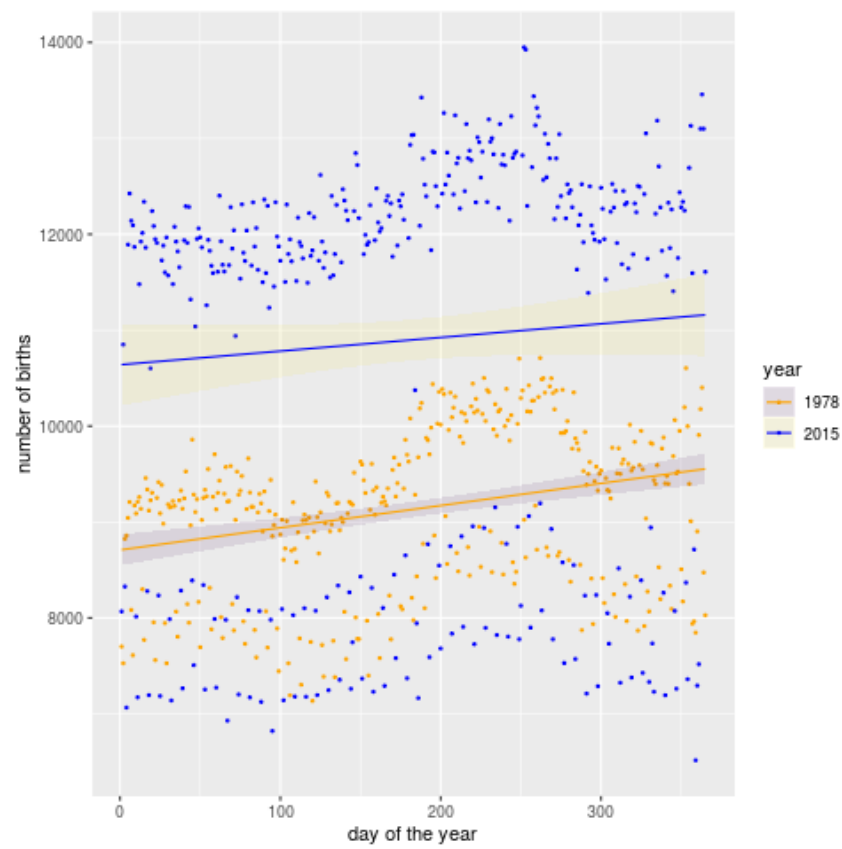


Figure 2: Graph from question 6

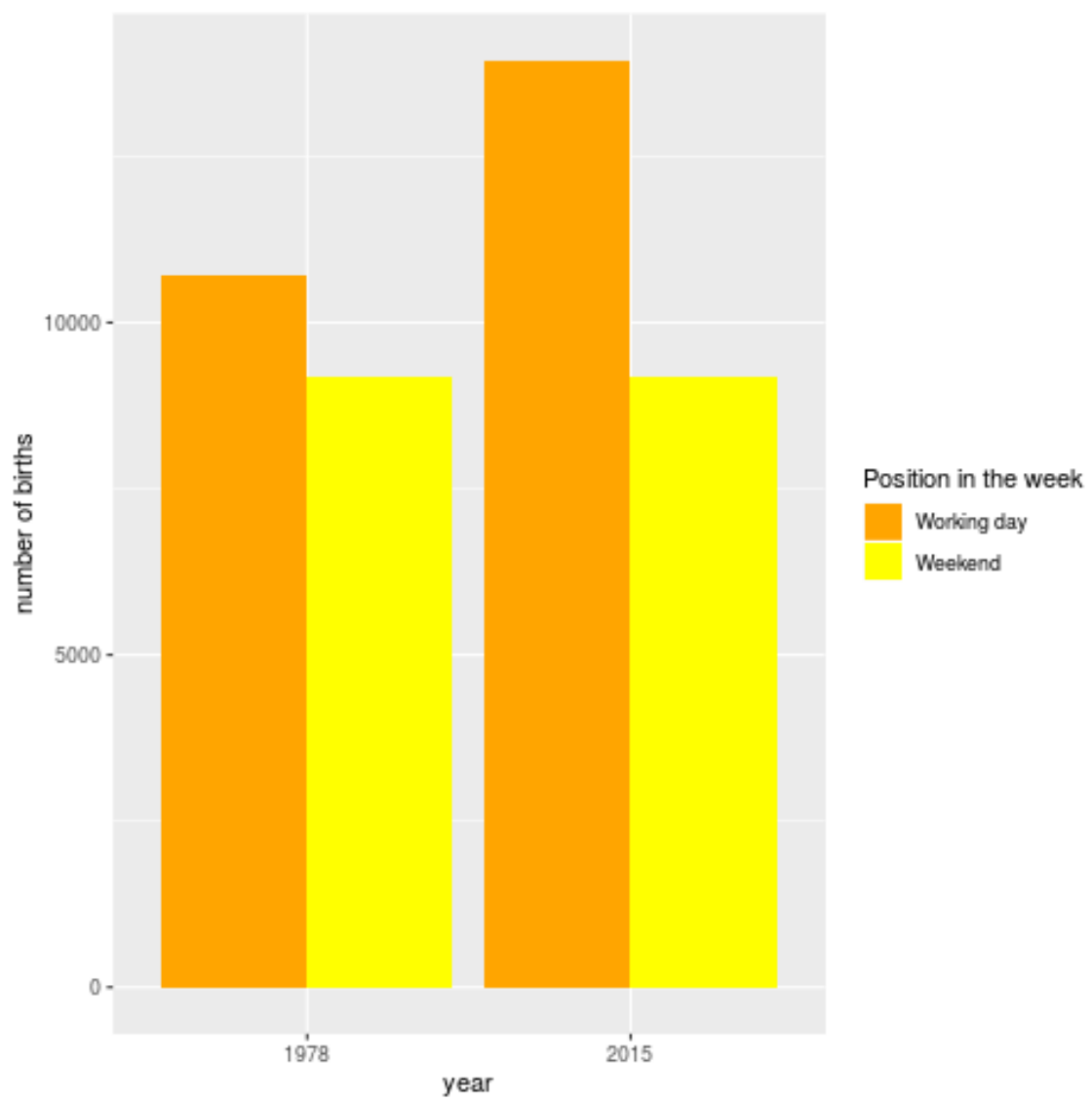


Figure 3: Graph from question 8

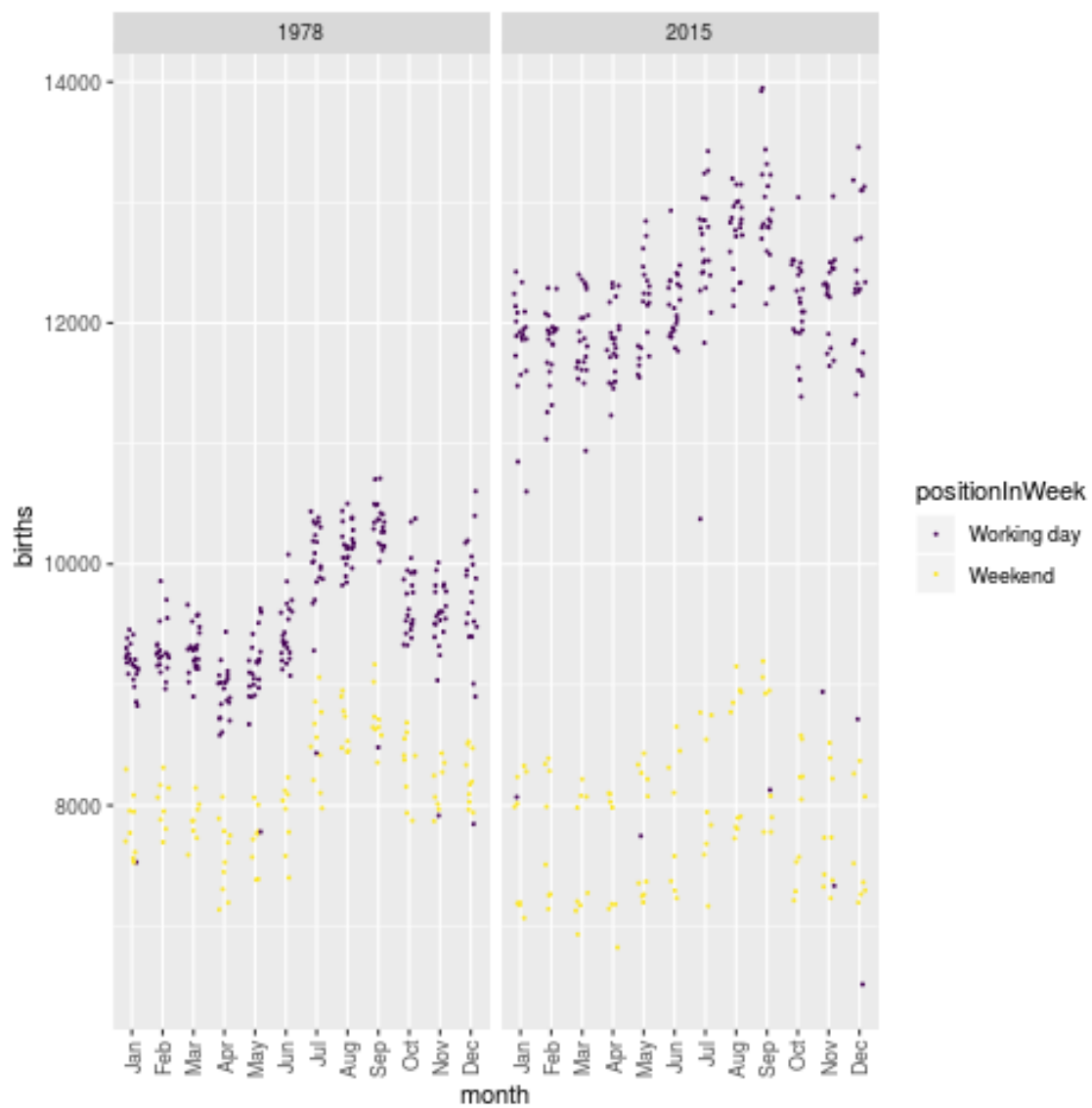


Figure 4: Graph from question 9