

# Lab 11

David Stenning

The folder `Lab11Data` contains several CSV data files.

```
dfiles <- dir("Lab11Data",full.names=TRUE)
dfiles
```

```
## [1] "Lab11Data/study1.csv" "Lab11Data/study2.csv" "Lab11Data/study3.csv"
## [4] "Lab11Data/study4.csv" "Lab11Data/study5.csv" "Lab11Data/study6.csv"
## [7] "Lab11Data/study7.csv" "Lab11Data/study8.csv" "Lab11Data/study9.csv"
```

1. Write R code to read in the first file. Print the tibble that you just read in. Use `names()` to change the column names of the tibble to `x` and `y`. Repeat for the second file. How many observations are in these first two files?
2. Use `vector()` to create an empty vector called `ff` that is of mode “list” and length 9. Now write a `for()` loop to loop over the 9 files in `dfiles` and for each (i) read the file in to a tibble, and change the column names to `x` and `y` as in part (1), and (ii) copy the tibble to an element of your list `ff`.
3. Write a function called `read.study_data` that takes a vector of data file names (like `dfiles`) as input, reads the data files into a list, assigns class “study\_data” to the list, and returns the list. Your function should use `length(dfiles)` to determine the number of files.
4. Write a function `plot.study_data()` that takes an object of class “study\_data” as input. The first 5 lines of your function should be the following, which creates a tibble with columns `study`, `x` and `y`:

```
dat <- NULL
for(i in seq_along(ff)) {
  d <- ff[[i]]
  dat <- rbind(dat,tibble(study=i,x=d$x,y=d$y))
}
```

Have your function coerce `study` to a factor, and then call `ggplot()` to make a plot of `y` *versus* `x`, with different colours for the different studies. Add points and smoothers to your plot.