# Spark and sparklyr setup

## Installing Java 8, sparklyr and Spark 2.4.5

The sparklyr package makes installing Spark and Scala really easy, but first we need to make sure that Java 8 is installed on your machine.

To be honest, Java versioning is a bit of a mess, and if you install multiple Java versions on your computer it can be difficult to select the version to use. So we recommend removing Java 11 if you find it on your system.

- **Check Java version**. In the RStudio console, run the command `system("java -version")`.
  - If you see nothing, an error, or a prompt to 'install a JDK', that's fine as it means no Java is currently installed - skip to the **Install Java 8** point.
  - If you see a Java version that starts with '1.8', that's fine too, as it means Java 8 is installed - skip to the **Install sparklyr** point.
  - If you see a Java version higher than '1.8' you need to uninstall the later version of Java - skip to the **Removing later Java versions** point.
- **Removing later Java versions**. In the Terminal `cd /Library/Java/JavaVirtualMachines` and have a look at the directories there (`ls`). These are the various versions of Java installed on your machine. You should remove any directories for Java versions after '1.8'. For example, if I found a directory `adoptopenjdk-11.jdk`, I would remove it with the command `sudo rm -rf adoptopenjdk-11.jdk` (be careful typing this command).
- **Install Java 8** In the Terminal, change to your home directory (`cd ~`) and then execute `brew cask install adoptopenjdk/openjdk/adoptopenjdk8` (it will update Homebrew and may ask for your password). Afterwards, back in the RStudio console, try `system("java -version")` again: hopefully you will now see a version starting '1.8'. Try restarting your R session and trying again if you see any other output.
- **Install sparklyr** Install the R sparklyr package in the normal way [e.g. in the RStudio console, run `install.packages("sparklyr")`].
- **Use sparklyr to install Spark** Next we'll use sparklyr to install the current stable version of Spark (2.4.5). So, in a code block, execute

```r
library(sparklyr)
spark_install(version = "2.4.5") # this is quite a large download
```

Once that's done, test your Spark installation with this code block

```r
sc <- spark_connect(master = "local")
spark_cars <- copy_to(sc, mtcars)
```

If these run fine you're good to go! **Any problems, we can help you fix them, don't worry!**