

Lesson 1: Introduction to the Spark Environment

1.7 Getting Set Up: R Kernel for Jupyter

Installation: IRKernel (Jupyter kernel for R)

- Make sure R is installed: <https://cran.r-project.org/bin/>
- Install kernel via R (get into an R shell):

```
install.packages(c('rvmq', 'repr', 'IRkernel', 'IRdisplay'),  
                repos = c('http://irkernel.github.io/', getOption('repos')))  
  
IRkernel::installspec()
```



Installation: IRKernel (Jupyter kernel for R)

And in the notebook:

```
# Set this to where Spark is installed
```

```
Sys.setenv(SPARK_HOME="/Users/jonathandinu/spark")
```

```
# This line loads SparkR from the installed directory
```

```
.libPaths(c(file.path(Sys.getenv("SPARK_HOME"), "R", "lib"), .libPaths()))
```

<https://github.com/apache/spark/tree/master/R>



Installation: IRKernel (Jupyter kernel for R)

And in the notebook: `library(SparkR)`

<https://github.com/apache/spark/tree/master/R>



Installation: IRKernel (Jupyter kernel for R)



<https://github.com/apache/spark/tree/master/R>



Note: If for any reason you cannot get Spark installed on your OS following these instructions, Cloudera and Hortonworks provide Linux VMs with Spark installed.

- http://www.cloudera.com/content/cloudera/en/downloads/quickstart_vms/cdh-5-4-x.html
- <http://hortonworks.com/products/hortonworks-sandbox/#install>



Review

- Command-line Spark shell: `./bin/pyspark`
- Spark module: `import pyspark as ps`
- Jupyter Notebook interface: `ipython notebook`
- Also R support in the notebook (or RStudio)!



Next Up: Spark Programming Basics

