

# Big Data con R Tidyverse y Spark

## 1.1. Intro slides

### 1.1.1. R & Big Data

<https://github.com/rstudio/webinars/blob/master/14-Work-with-big-data/14-Work-with-big-data.pdf><sup>[1]</sup>

### 1.1.2. Tidyverse

- <https://github.com/rstudio/webinars/blob/master/46-tidyverse-visualisation-and-manipulation-basics/00-Tidyverse-webinar.pdf><sup>[2]</sup>
- <https://github.com/rstudio/webinars/blob/master/55-ciencia-de-datos-R/ciencia-de-datos-R.pdf><sup>[3]</sup>

### 1.1.3. Spark (sparklyr):

<https://github.com/rstudio/webinars/blob/master/42-Introduction%20to%20sparklyr/Introducing%20sparklyr%20-%20Webinar.pdf><sup>[4]</sup>

### 1.1.4. Speeding up Spark via R via Arrow

<https://arrow.apache.org/blog/2019/01/25/r-spark-improvements/><sup>[5]</sup>

## 1.2. Intro Exercise

See:

<https://gitlab.com/radup/curs-r-introduccio/blob/master/codi/extra.tips.bigdata.R><sup>[6]</sup>

## 1.3. References:

- TheRinSpark Book  
<https://therinspark.com/><sup>[7]</sup>
- RStudio Webinar: Introducing an R interface for Apache Spark  
<https://www.rstudio.com/resources/webinars/introducing-an-r-interface-for-apache-spark/><sup>[8]</sup>  
<https://github.com/rstudio/webinars/blob/master/42-Introduction%20to%20sparklyr/sparklyr-webinar1.Rmd><sup>[9]</sup>
- Some online tutorials
  - 30Gb DataSet
  - Text mining using sparklyr

- Some cheatsheets:

<https://www.rstudio.com/resources/cheatsheets/><sup>[10]</sup>

- And some in Spanish:

<https://www.rstudio.com/resources/cheatsheets/><sup>[11]</sup> > Spanish Translations – Traducciones en español

---

<sup>[1]</sup> <https://github.com/rstudio/webinars/blob/master/14-Work-with-big-data/14-Work-with-big-data.pdf>

<sup>[2]</sup> <https://github.com/rstudio/webinars/blob/master/46-tidyverse-visualisation-and-manipulation-basics/00-Tidyverse-webinar.pdf>

<sup>[3]</sup> <https://github.com/rstudio/webinars/blob/master/55-ciencia-de-datos-R/ciencia-de-datos-R.pdf>

<sup>[4]</sup> <https://github.com/rstudio/webinars/blob/master/42-Introduction%20to%20sparklyr/Introducing%20sparklyr%20-%20Webinar.pdf>

<sup>[5]</sup> <https://arrow.apache.org/blog/2019/01/25/r-spark-improvements/>

<sup>[6]</sup> <https://gitlab.com/radup/curs-r-introduccio/blob/master/codi/extra.tips.bigdata.R>

<sup>[7]</sup> <https://therinspark.com/>

<sup>[8]</sup> <https://www.rstudio.com/resources/webinars/introducing-an-r-interface-for-apache-spark/>

<sup>[9]</sup> <https://github.com/rstudio/webinars/blob/master/42-Introduction%20to%20sparklyr/sparklyr-webinar1.Rmd>

<sup>[10]</sup> <https://www.rstudio.com/resources/cheatsheets/>

<sup>[11]</sup> <https://www.rstudio.com/resources/cheatsheets/>