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[1]

1.1.2. Tidyverse

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

1.1.3. Spark (sparklyr):

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

1.1.4. Speeding up Spark via R via Arrow

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

1.2. Intro Exercise

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

1.3. References:

- TheRinSpark Book
  https://therinspark.com/[7]

- RStudio Webinar: Introducing an R interface for Apache Spark
  https://www.rstudio.com/resources/webinars/introducing-an-r-interface-for-apache-spark/[8]
  https://github.com/rstudio/webinars/blob/master/42-Introduction%20to%20sparklyr/sparklyr-webinar1.Rmd[9]

- Some online tutorials
  - 30Gb DataSet
  - Text mining using sparklyr

- Some cheatsheets: https://www.rstudio.com/resources/cheatsheets/[10]
    - And some in Spanish: https://www.rstudio.com/resources/cheatsheets/[11] > Spanish Translations – Traducciones en español

---

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

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

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

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

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

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

[7] https://therinspark.com/

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

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

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

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