References

Books

R for Data Science: Import, Tidy, Transform, Visualize, and Model Data
http://a.co/fzpONDS

Report Writing For Data Science
https://leanpub.com/reportwriting

Data Science SaaS

Apache SystemML: https://systemml.apache.org/

Google Cloud Lab: https://cloud.google.com/datalab/

Articles

Running Jupyter Notebook on Google Cloud Platform in 15 min
https://medium.com/towards-data-science/running-jupyter-notebook-in-google-cloud-platform-in-15-min-61e16da34d52

GitHub as a Fast Track to Interviews
https://medium.com/@bfil/github-as-a-fast-track-to-interviews-2cdf3198eb2f

Introduction to AWS Billing and Cost Management Tools
https://cloudacademy.com/blog/introduction-to-aws-billing-and-cost-management-tools-part-1-of-3/

https://cloudacademy.com/blog/introduction-to-aws-billing-and-cost-management-tools-part-2-of-3/

https://cloudacademy.com/blog/introduction-to-aws-billing-and-cost-management-tools-part-3-of-3/

https://aws.amazon.com/s3/pricing/

Opinionated Analysis Development
https://www.rstudio.com/resources/videos/opinionated-analysis-development/

Where Trumponomics Is Working
https://www.washingtontimes.com/news/2017/oct/22/trump-tax-cuts-will-work-if-given-time/

Online Courses

Introduction to Sparklyr for Data Science:
https://www.safaribooksonline.com/library/view/introduction-to-sparklyr/9781491996508/

Learning Path: Jupyter Notebook for Data Science Teams:
https://www.safaribooksonline.com/learning-paths/learning-path-jupyter/9781491995648

Reproducible Research:
https://www.coursera.org/learn/reproducible-research

Git Hub Repos

gradle-plugin-r: https://github.com/jamiefolson/gradle-plugin-r

This is a gradle plugin for building and installing R packages. It also supports roxygen2, which is run prior to building the package.

Sparkmagic: https://github.com/jupyter-incubator/sparkmagic

Sparkmagic is a set of tools for interactively working with remote Spark clusters through Livy, a Spark REST server, in Jupyter notebooks. The Sparkmagic project includes a set of magics for interactively running Spark code in multiple languages, as well as some kernels that you can use to turn Jupyter into an integrated Spark environment.

sparkDemos: https://github.com/rstudio/sparkDemos

dbplot: https://github.com/edgararuiz/dbplot

Do the calculations for ggplot in the Spark cluster instead of R memory.

Tools

Nodebook
http://multithreaded.stitchfix.com/blog/2017/07/26/nodebook/

Workflowr
https://jdblischak.github.io/workflowr/index.html

Software Carpentry
https://software-carpentry.org/

Test Driven Data Analysis
http://www.tdda.info/