**Universidad Nacional Agraria La Molina**

**Facultad de Agronomía**

Benites Alfaro, Omar ([obacc07@gmail.com](mailto:obacc07@gmail.com));[o.benites@cgiar.org](mailto:o.benites@cgiar.org)(1)

Gutierrez Vilchez, Pablo ([pgutierrez@lamolina.edu.pe](mailto:pgutierrez@lamolina.edu.pe))(2).

1. Departamento de Investigación Informática para Mejoramiento Genético, CIP, Lima.
2. Facultad de Agronomía - Departamento de Suelos - UNALM.

**Datos generales**

**Programa :** Proyecto de Agro-informática: Paquete de datos abiertos en R:

**Cropdatape**

**Requisito :** Excel Intermedio y Estadistica Basica

**Duración :** 10 semanas

**Vacantes :** 3 o 4 alumnos de pregrado

**Fecha : -**

**Requerimiento :** Laptop

**Justificación**

Actualmente una gran cantidad investigadores a nivel mundial incluyen en sus flujos de trabajo o *workflows*, herramientas computacionales para el acceso, análisis y reproducción de datos en sus investigaciones. Sin embargo, en Perú, son muy pocas las iniciativas que incorporan estas metodologías. El proyecto de agroinformatica-unalm trata de incorporar un nuevo paradigma de investigación, el cual trata de responder a la demanda de ciencia de datos en agricultura. Para ello, deseamos conformar un equipo de estudiantes, profesores e investigadores, los cuales mediante datos, métodos estadísticos, código libre, y la informática tratamos de resolver los diversos problemas durante colección, estandarización, análisis y fortalecer la toma de decisiones en el nuevo contexto de la agricultura inteligente.

En este proyecto se harán uso de herramientas basados en código libre, *lenguaje* ***R***, con el cual desarrollaremos el primer conjunto de datos abiertos u ***open access*** de diversos cultivos peruanos. Este conjunto de datos busca facilitar el libre acceso a la información de producción agrícola en el Perú a investigadores, catedráticos y estudiantes.

**Objetivos**

* Aprender el uso de herramientas para el análisis exploratorio de datos con el software R
* Desarrollo de gráficos estadísticos en R.
* Aprender la estandarización y manipulación de datos usando el software R.
* Creación de paquetes para el lenguaje R.
* Generación de una base datos abierta usando los datos estadísticas agrarias del Ministerio de Agricultura y Riego (MINAGRI).
* **Publicación del paquete** en el The Comprehensive **R** Archive Network (CRAN).
* Generar un círculo/grupo de investigación de Agroinformatica/Bioinformática de la Universidad Agraria (UNALM).

**Cronograma y contenido de curso**

|  |  |  |  |
| --- | --- | --- | --- |
| **Fecha** | **Horario** | **Horas** | **Contenido del proyecto** |
| **Semana 1** | **17:10- 18:40** | **1h 30min** | **Presentación y requisitos del proyecto** |
| **Lenguaje R, uso de librerías y fundamentos de agroinformatica en R** |
| **Programación en R: estructura de datos, tipos de variables e implementacion de codigo.** |
| **Importación de base de datos a R** |
| **Semana 2** | **17:10- 18:40** | **1h 30min** | **Manipulación de datos en R (dplyr, tidyr, doBy)** |
| **Resumen de base de datos en R** |
| **Visualización de datos en R (ggplot2)** |
| **Semana 3** | **17:10- 18:40** | **1h 30min** | **Control de versiones y Trabajo Colaborativo con Git, Github y Rstudio** |
| **Creación del Paquete-R de datos : Cropdatape** |
| **Compilación del Paquete en R y Rstudio usando devtools, roxygen y testthat** |
| **Semana 4** | **17:10- 18:40** | **1h 30min** | **Fundamentos de Data Management: Aplicación al MINAGRI data en R part1** |
|
|
| **Semana 5** | **17:10- 18:40** | **1h 30min** | **Estandarización de la Información: Aplicación al MINAGRI data en R part2** |
|
|
| **Semana 6** | **17:10- 18:40** | **1h 30min** | **Creación de la base de datos de cultivos del Minagri en R: Cropdatape dataset parte 1** |
|
|
| **Semana 7** | **17:10- 18:40** | **1h 30min** | **Creación de la base de datos de cultivos del Minagri en R: Cropdatape dataset parte 2** |
|
|
| **Semana 8** | **17:10- 18:40** | **1h 30min** | **Creación de la base de datos de cultivos del Minagri en R: Cropdatape dataset parte 3** |
|
|
| **Semana 8** | **17:10- 18:40** | **1h 30min** | **Creación de la base de datos de cultivos del Minagri en R: Cropdatape dataset parte 4** |
|
|
| **Semana 9** | **17:10- 18:40** | **1h 30min** | **Versión Previa del Paquete "Cropdatape"** |
|
|

**Softwares**

* R Cran
* R Studio
* Excel

**Requerimiento técnico del salon**

* Conexión a internet
* Enchufes para cargar laptops
* Proyector Multimedia
* Pizarra

**Referencias**

* Hadley Wickhman. ggplot2: Elegant Graphics for Data Analysis (Use R!) 2nd ed. 2016 Edition.
* Norman Matloff. The Art of R Programming: A Tour of Statistical Software Design. Edition 2014
* [J. Crawley](https://www.amazon.es/s/ref=dp_byline_sr_book_1?ie=UTF8&field-author=Michael+J.+Crawley&search-alias=stripbooks). The R Book. Edition 2012.
* Hadley Wickhman. R Packages 1st Edition. 2014
* Chang, W. 2013. R Graphics Cookbook. Eds. M Loukides; N Courtney. First Edit United State of America, O´Reilly Media, 413.
* De Mendiburu, F. 2007. Análisis Estadístico con R. 2007: 66.
* De Mendiburu, F. 2016. agricolae: Statistical Procedures for Agricultural Research. Disponible en <http://cran.r-project.org/package=agricolae> (R CRAN).
* Mendiburu, F de. 2014. Tutorial de agricolae. 2014: 84.
* Kabacoff, RI. 2011. R IN ACTION: Data analysis and graphics with R. Ed. S Stirling. First Shelter Island, NY, MANNING, 474.
* Paradis, E. 2003. R para Principiantes. 2003: 61.
* Ihaka, R; Gentleman, R. 1996. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics feb. 1996.
* Nuzzo, R. 2015. How scientists fool themselves – and how they can stop. Nature 526(7572): 182-185.