Introduction to R

Module 02-02, R for Data Science

Billy Fetzner

Why do data scientists love R?

Free and open source

Effects of R Being Free and Open-Source

- How they are taught

- Cheap option

Excellent Performance

- Easy to learn

- Perform calculations

- Statistical Analysis

Great Packages

- **Cleaning data**
  - dplyr
  - Readr
  - Tidyverse

- **Visualizing Data**
  - Ggplot2

- **Machine Learning**
  - MICE (missing values)
  - Rpart & PARTY (data partitions)
  - CARET (classification and regression)
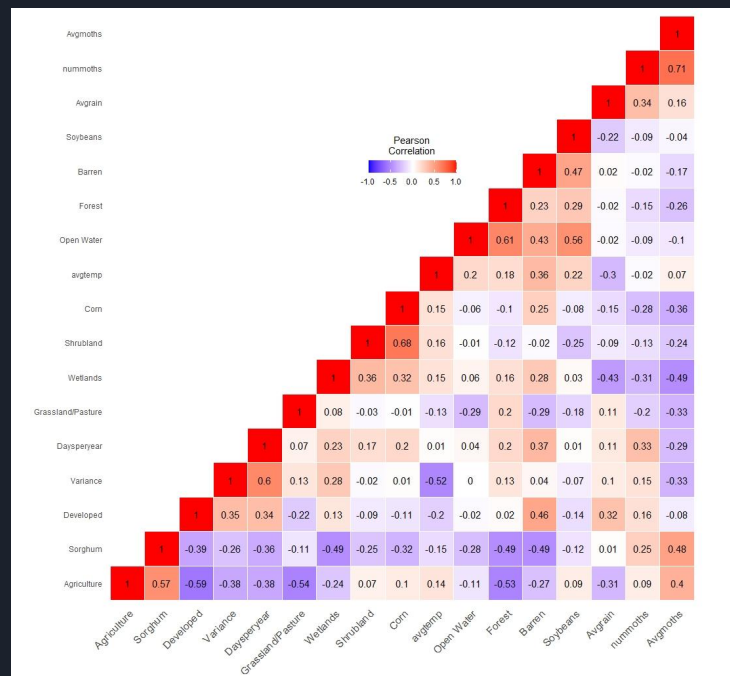 - randomFOREST (creating decision trees)

Figure created by author

Online Forums

- High popularity

- Easy communication

Python vs. R

- Python AND R

- Python is better if you have a programming background

- R is better for beginners

Summary

- Free and open-source

- Great performance

- Packages

- Online community