R for Raster, (B)ridge for Big Data: What is new in R-ArcGIS bridge?

Marjean Pobuda & Orhun Aydin
Presentation Slides & Materials

• https://spatialstats.github.io/presentations/
Introducing the R-ArcGIS Bridge
Why R?

• A language designed with data analysis in mind
  - Powerful core data structures and operations
• Unparalleled breadth of statistical routines
  - Both cutting-edge research and field specific methods
• CRAN: Over 12,000 available packages
• Versatile and powerful plotting
R Objects

Data Frames

- Treats tabular (and multi-dimensional) data as labeled, indexed series of observations.

```r
# Create a data frame out of an existing source
df.from.csv <- read.csv(
  "data/growth.csv",
  header=TRUE)

# Create a data frame from scratch
quarter <- c(2, 3, 1)
person <- c("Goodchild",
            "Tobler",
            "Krige")
met.quota <- c(TRUE, FALSE, TRUE)
df <- data.frame(person,
                 met.quota,
                 quarter)
```
R Objects
Spatial Data Types

- Objects designed to contain attribute information, like a data frame, but also contains a spatial component.
The R-ArcGIS Bridge

- Analysts can tap into novel or field-specific statistical methods.
- ArcGIS developers can create tools and toolboxes that integrate ArcGIS and R.
- ArcGIS users can access R code through geoprocessing scripts in ArcMap, Pro, or Online.
- R users can access organizations GIS data and take advantage of improved functionality.

https://r-arcgis.github.io/
The R-ArcGIS Bridge

- Store your data in ArcGIS
- Quickly access it in R
- Returns R objects back to ArcGIS native data types
- Knows how to convert data back and forth from ArcGIS to R and vice versa

https://r-arcgis.github.io/assets/arcgisbinding-vignette.html
Requirements for the R-ArcGIS Bridge

**ArcGIS Pro**
- R 3.2.2 (or later)
- ArcGIS Pro 1.1 (or later)

**ArcMap**
- ArcMap 10.3.1 (or later)
- RStudio Optional
Installing the R-ArcGIS Bridge

ArcGIS Pro

Project > Options > Geoprocessing

R workspace (RStudio)

1) Load R-ArcGIS bridge library
2) Connect to ArcGIS

library(arcgisbinding)
arc.check_product()
Using the R-ArcGIS Bridge

1) Open ArcGIS data, tables, layers

   gis_data <- arc.open(path = 'C:/Data/Seagrass.shp')

2) Load dataset to R data frame

   R_data <- arc.select(gis_data, fields, SQL, spatial ref)
Demo

Getting Started
Introducing Raster Support
New in the **arcgisbinding** Package

**Raster Support**

- Ability to read and write raster data
  - Handle big data raster data with the ability to read in chunks by bands
  - Compatibility with CRF format and Mosaic Datasets
- Customize selections and subsets
  - Create subsets by bands or pixel rows and columns
  - Resample options available
  - Select desired pixel format for specific analyses
New in the **arcgisbinding** Package

Additional Features

- Compatibility with `sf` package
- Maintain spatial geometries when working with `dplyr`
- Ability to delete objects written to ArcGIS from R
Using the R-ArcGIS Bridge

**Raster Data**

1) **Open ArcGIS single band or multiband rasters:**
   
   ```r
   gis_data <- arc.open(path = 'C:/Data/MyRaster.img')
   ```

2) **Customize data details:**
   
   ```r
   Arc_data <- arc.raster(gis_data, nrow, ncol, bands, extent, pixel_type, resample_type)
   ```
Using the R-ArcGIS Bridge

Raster Data

3) Convert ArcGIS data to desired R object

R_data <- as.raster(Arc_data) or ... as.<insert desired type>
Demo
Raster Support
ArcGIS Pro - Microsoft R Integration
Using the R-ArcGIS Bridge with Microsoft R

Microsoft Open R

• Microsoft Open R is a publicly available R-version

• Contains almost all CRAN libraries
  • It lags CRAN in functionality, Microsoft follows CRAN releases to update

• Provides integration to other Microsoft tools such as R-Server
Using the R-ArcGIS Bridge with Microsoft R
ArcGIS Bridge- Microsoft R Connection

• Connection to `argisbinding` package is `same` as CRAN version of R

• Can be used as the background R version within ArcGIS Pro

• Usage from Pro is exactly the same as CRAN
Comparing Different R Distributions

- Allows multi-thread parallelization
- Allows parallelization and remote computing
- Processing limited to 2 threads
- All processing is handled locally
- Microsoft’s implementation of R
  - Publicly available
  - Efficient matrix operations
- Open-source
  - Contains newest libraries

Microsoft R Server
Microsoft R Client
Microsoft R Open
CRAN
How is Microsoft R useful?
Time Series Rasters and Mosaics

Raster data can become a big data problem, quickly

**Mosaics**: Data structure to store/process rasters in space and time
Working with Mosaic Datasets

1) Open mosaic data with `arc.open`

   ```r
   open_obj <- arc.open(path = 'C:/Data/Seagrass.shp')
   ```

2) Use `arc.raster` to access mosaic info and operations

   ```r
   Mos <- arc.raster(open_obj, list(method = methodType, operator = operatorType)
   ```

Method belongs to a family of sorting rules used by ArcGIS Pro
Operator is the operator used to Resolve Overlaps in ArcGIS Pro
Demo

Using Microsoft Open R with Rasters and Mosaics

```r
## IMPORT LIBRARIES FOR ANALYSIS
library(arclibaries)
library(spatstat)
library(raster, quietly = TRUE, warn.conflicts = FALSE)

## INITIALIZE CONNECTION WITH ARGIS PRO
arc.check_product()

## READ IN THE MOSAIC DATASET
open_obj <- arc.open('C:/Users/owa8849/Desktop/DevSummit2018/R_Bridge/BridgeDemoMosaics/Mosaics.gdb/Local_T_Model')

## DEFINE THE MULTI-BAND RASTER AS A RASTER OBJECT
ras_obj <- arc.raster(open_obj, mosaic = list(method = "NorthWest", operator = "Sum"))

## VISUALIZE THE MOSAIC RASTER
plot(as.raster(ras_obj))

## CHANGE THE SCALE FOR SIZE CONSIDERATIONS
ras_obj <- arc.raster(open_obj, mosaic = list(method = "NorthWest", operator = "Sum"), nrow = 1000, ncol = 1000)
plot(as.raster(ras_obj))
```
Matrix Multiplication Benchmark on Rasters
Microsoft Open R

n-by-n Raster Multiplication Benchmark

Run Time (sec)

Raster Dimension

CRAN
Microsoft R
Using the R-ArcGIS Bridge
Microsoft Open R

- Image convolutions are matrix multiplications
- Window-based operations and image operators speed up drastically
- Integrates to bigger data platforms of Microsoft such as Azure and R-Server
R-Server Integration
Microsoft Open R

- R-Server allows running jobs on multiple compute nodes synchronously, asynchronously
- Backbone for scaling up large data operations in R
- Specific R libraries developed by Microsoft and RevoR allow parallelization
Demo

Using Microsoft Open R with Rasters and Mosaics
Training courses
www.esri.com/training

*Using the R-ArcGIS Bridge*

*Integrating R Scripts into ArcGIS Geoprocessing Tools*

*Analyze Crime Using Statistics and the R-ArcGIS Bridge*

Tutorials, blogs, and other resources

https://r-arcgis.github.io

https://r-arcgis.github.io/assets/arcgisbinding-vignette.html

https://geonet.esri.com/groups/rstats/activity

https://learn.arcgis.com/en/