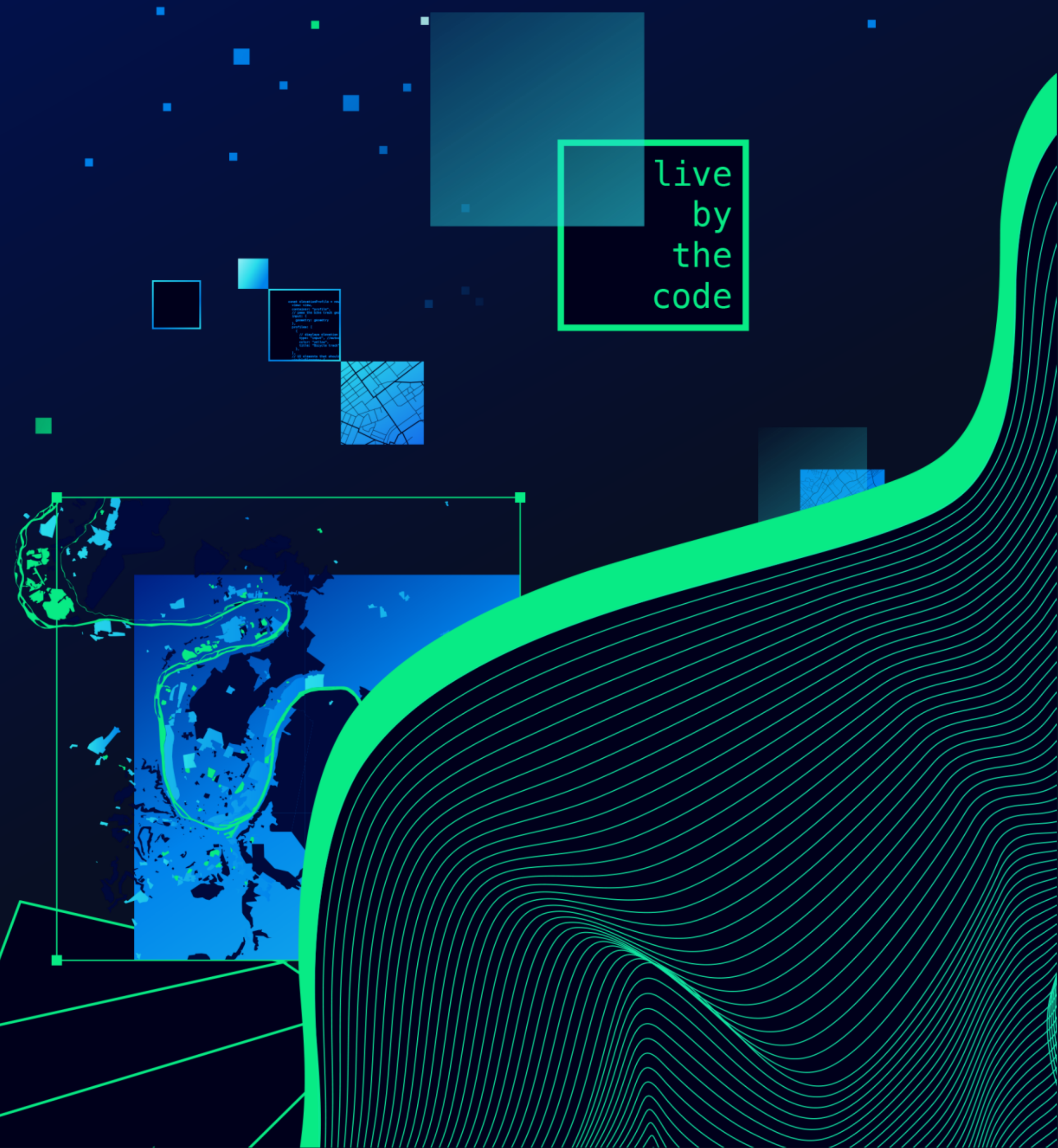
Doing Spatial Data Science Using R and ArcGIS

Orhun Aydin, PhD

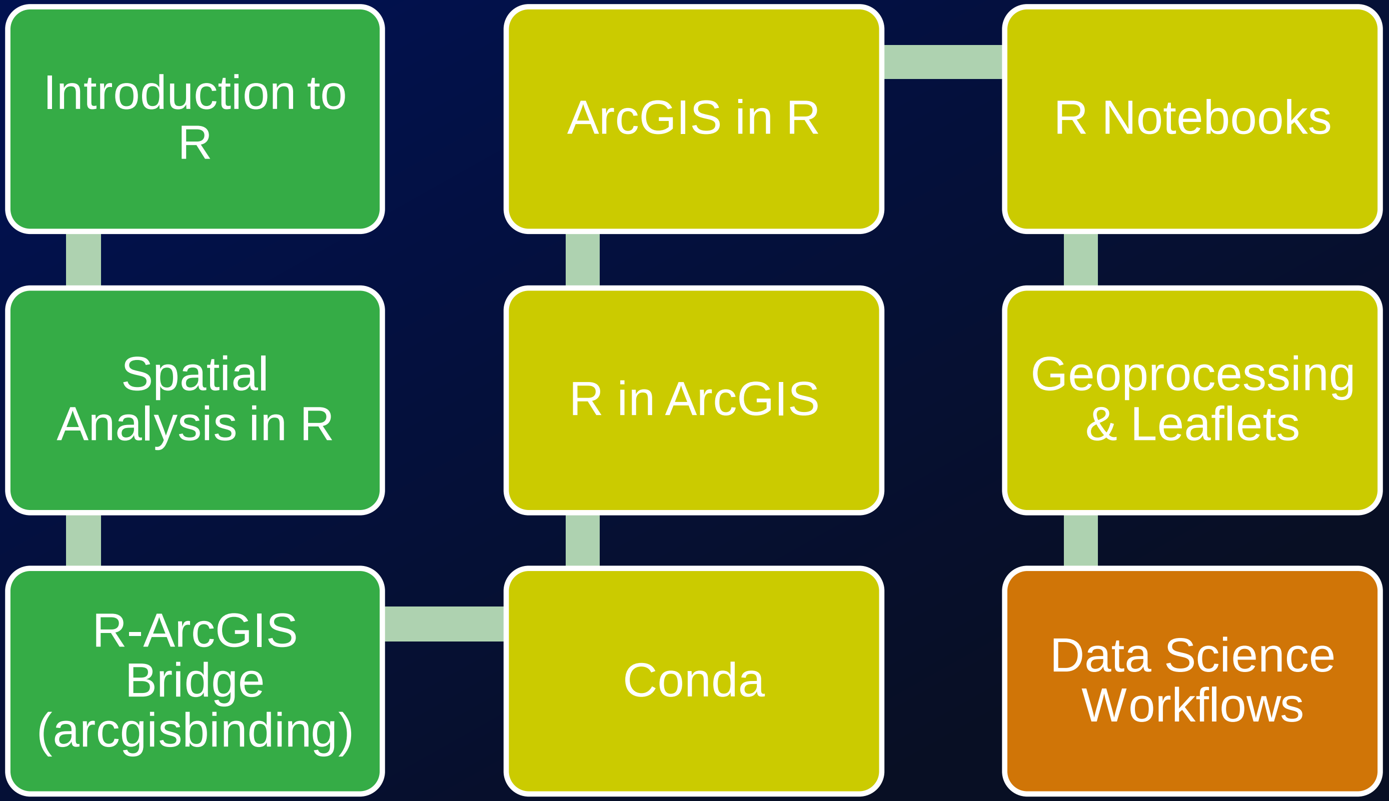
2022 ESRI DEVELOPER SUMMIT

Seminar Overview

Today’s Schedule



Course in a Nutshell



Introduction to R

Spatial Analysis in R

R-ArcGIS Bridge (arcgisbinding)

Conda

R in ArcGIS

ArcGIS in R

R Notebooks

Geoprocessing & Leaflets

Data Science Workflows

Workshop Schedule - Morning

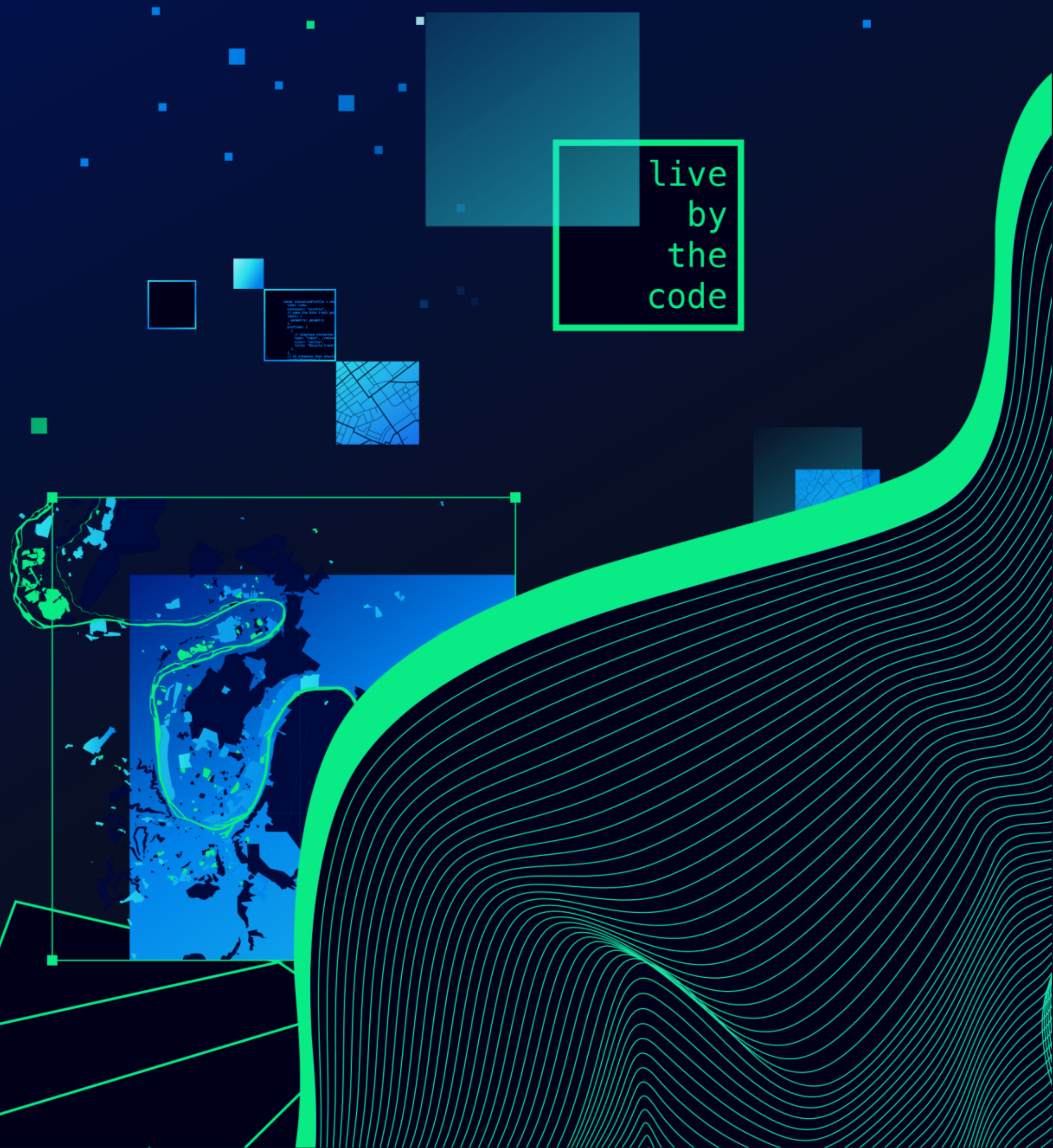
Section Title	Details	Time
1. Introduction to R Language	R Language R Dataframe R Studio	8:00 am – 8:30 am
2. Spatial Data Science in R	Spatial data representations Sf, sp raster	8:30 am – 9:00 am
3. The R-ArcGIS Bridge	Setting Up Spatial data I/O Manipulating data	9:00 am – 10:00 am
Break		10:00 am – 10:15 am
4. R Script Tools	Hello World Wrapping R functionality	10:15 am – 11:30 am
5. Getting Started with Conda	Setting Up r-arcgis-essentials	11:30 am – 12:00pm
Lunch Break		12:00 pm – 1:00 pm

Workshop Schedule - Afternoon

Section Title	Details	Time
5. Getting Started with R Notebooks	Starting an R Notebook Markdown/Code	1:00 pm – 2:30 pm
Break		2:30 pm – 2:45 pm
6. Geoprocessing and interactive maps	reticulate esri.leaflet	2:45 pm – 3:45 pm
7. Data Science Workflows	Hands-On Problem Solving	3:45 pm – 4:30 pm
8. Q&A and Resources		4:30 pm – 5:00 pm

The R Language

An Overview



Why Use R?

- R is a free and open-source programming 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 12,700 available packages (doubled since 2016!)
- Versatile and powerful plotting

Language features: core types

- Data types you’re used to seeing:
  - `Numeric - Integer - Character - Logical - timestamp`
- But others that you may not be:
  - `vector - matrix - data.frame - factor`

Language features: data frames

- Treats tabular (and multi-dimensional) data as a *labeled*, *indexed* series of observations. Sounds simple, but is a game changer over typical software which is just doing 2D layout (e.g. Excel)

- This concept of the data frame is core, and matches nicely to how data is manipulated and managed in geographic contexts

R Resources

- A strong and engaged user community
  - Easy to find tutorials and resources for help
    - Cross-Validated - https://stats.stackexchange.com/
    - R-bloggers - https://www.r-bloggers.com/
    - Kaggle - https://www.kaggle.com/competitions
    - Hadley Wickham’s Advanced R - http://adv-r.had.co.nz/
    - Introduction to Statistical Learning with Applications in R - http://www-bcf.usc.edu/~gareth/ISL/
    - Applied Spatial Data Analysis with R - http://www.asdar-book.org/
  - A variety of active user groups
    - R Consortium - https://www.r-consortium.org/
    - Meetup Communities – https://www.meetup.com/find/tech/
    - R-Ladies - https://rladies.org/

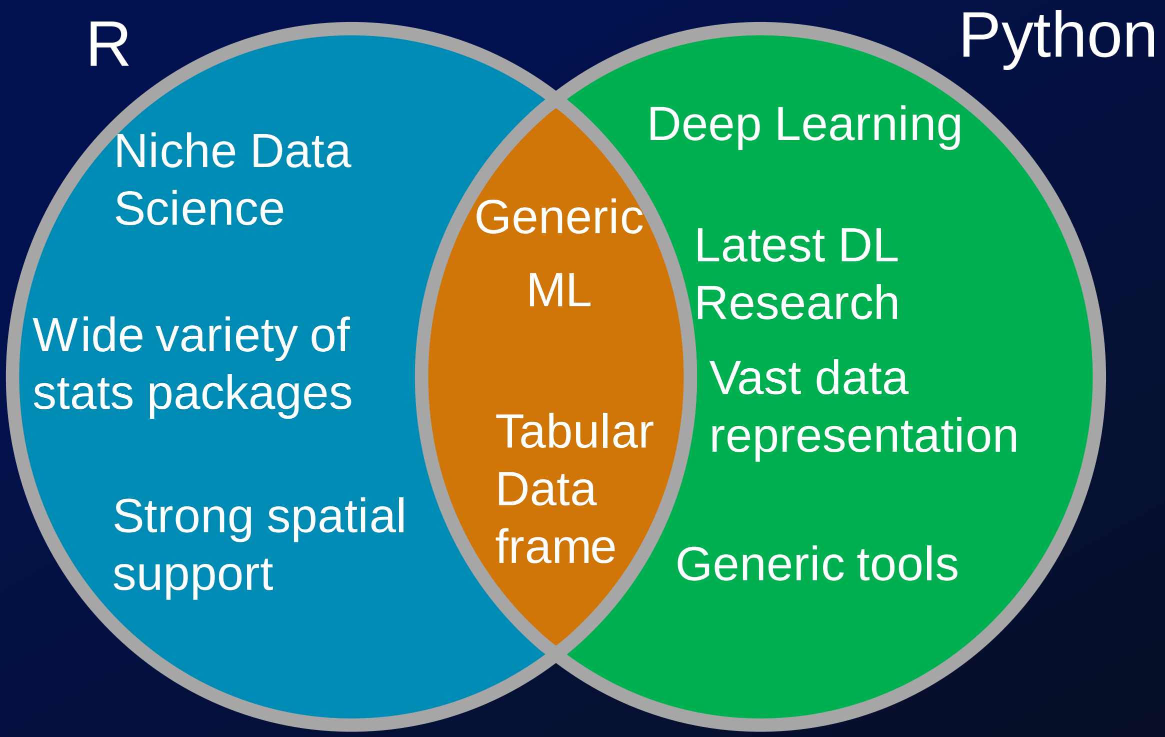
Strengths of the R Language

- Breadth of available geospatial analysis packages
    - 16797 packages as of December 2020
- Niche data science functionality
    - Spatial data science
    - Biostatistics
- Ease of programming
- Ecosystem built for open-science

Weaknesses of R Language

- Performance
  - R is considerably slower than Python, about 5-10 times
  - Loops are notoriously slow
- Memory
  - Memory bottlenecks occurs very frequently with medium size (1, 2 GB) data
  - Inefficient R code is not as forgiving as Python
- Security (will not be an issue addressed in this course)

R and Python for Geocomputation



- R serves vast amounts of field-specific data science packages
  - Biostatistics
  - Geostatistics
  - Econometrics
- Python has general purpose data-science libraries
  - Deep Learning (Tensorflow, …)
  - Machine Learning (scikit-learn, …)
  - Used for analysis & build scalable software

How to work with R?

RStudio

- IDE speficially designed for R
- Easiest way to get started
- Free
- Integrated Visualization and Variable Space
- Support for R Markdown and & notebooks

jupyter

- Native support for R notebooks
- Interactive analysis
- Free
- Setup takes some effort

Vim

- Editors that can be pointed to an R installation
- Vim and Visual Code are free
- Cumbersome to setup for non-developers
- Text editor with syntax marking

Installing Packages

- Easiest method to install packages
- install.packages(“<package_name>”)
- For packages that do not exist in CRAN (main repository for R packages)

  Download the R package repo

  Extract to Folder containing packages
- Devtools
  - Tool used for building a package that require compilation
-

Defining R Variables

- var <- my_fun(input1, input2)
- Variable var contains the output from my_fun

Acceptable Variable Names	Unacceptable Variable Names
var1	var$1
var_1	var@1
var.1	1.Var
	var!1

R Data Types

- Character – Stores str
- Complex – Stores complex numbers. *5 + 4i*
- Integer – Stores integers, numbers without floating points. *2, 5, 100*
- Logical – Stores Boolean variables. *True, False*
- Numeric – Generic container for any number. *2, 3.14*

R Data Structures

- Data Frame
- Factor
- List
- Matrix
- Vector (Atomic Vector)

More R Data Structures

- **Homogeneous (Single-Type) Data**
- Can store one type of data
  - Array(n-dimensional)
  - Matrix (2D)
  - Vector (1D)

- **Heterogeneous (Multi-Type) Data**
- Can store multiple types of data
  - List
    - r_list <- (‘geo’, 1.2, ‘e’, obj)
  - Data Frame

Time	Lat	Lon	Type
1/1/90	44.55	72.58	Built
1/1/10	44.55	72.58	Water

Summarizing Data Structures

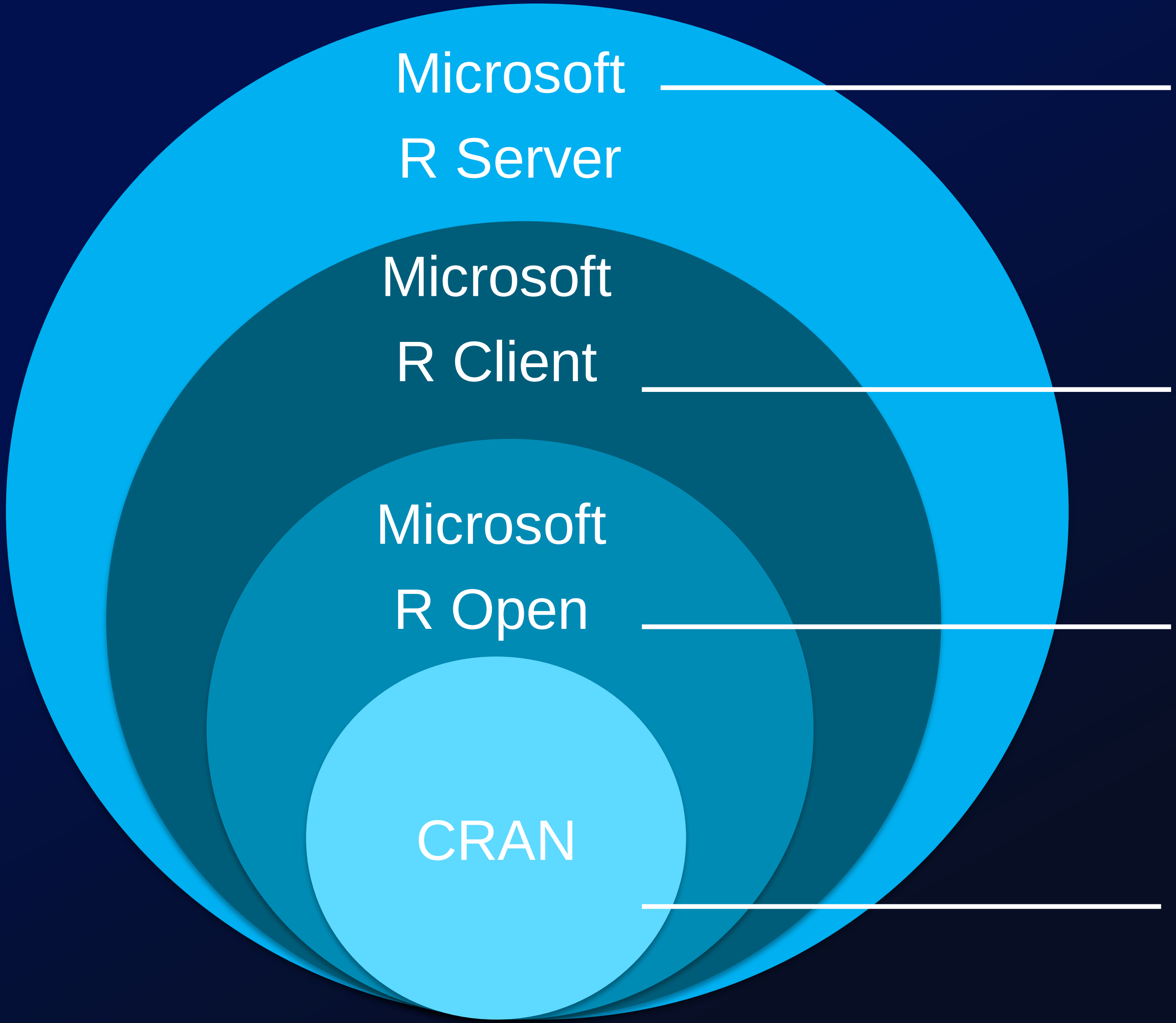
- Common functions to examine variable, var
- typeof(var)
- length(var)
- class(var)
- dim(var)

Getting Help from Documentation

- I forgot what this function did
- I do not remember what kind of inputs I need
- Getting help

```
? <name of the function>
? arc.open
```

Different Versions of R



Allows multi-thread parallelization

Parallelization

Processing limited to 2 threads

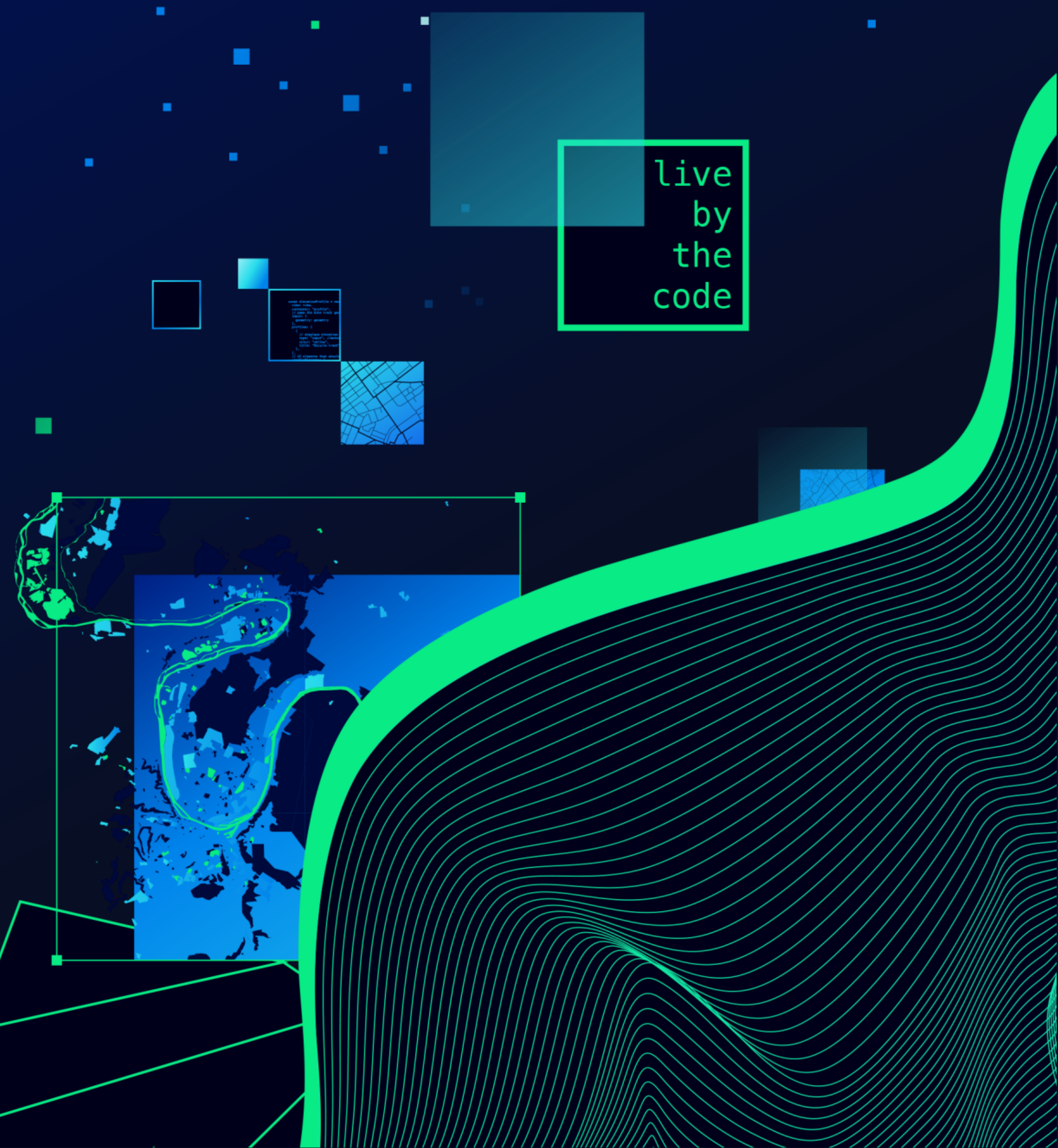
All processing is handled locally

Microsoft’s implementation of R
Publicly available
Efficient matrix operations

Open-source
Contains newest libraries

Spatial Data in R

Packages and the Ecosystem



Spatial R Data Types

- Vector Data
  - sf, sp packages
  - Represent spatial information efficiently
  - Accepted by most packages that require spatial dataframes
- Raster Data
  - raster package

☢ Is this the same as vector data structure from the previous slide?

Spatial Vector Data Representation in R

- Same pattern: Geometry Info + Attributes
- sf and sp only differ in representing geometry information
- Once vector data is in sf and sp format
  - Spatial operators: join, dissolve, merge, etc.
  - Reprojecting the data (change projection system)
  - Spatially subset the data

Spatial Vectors in R - sf

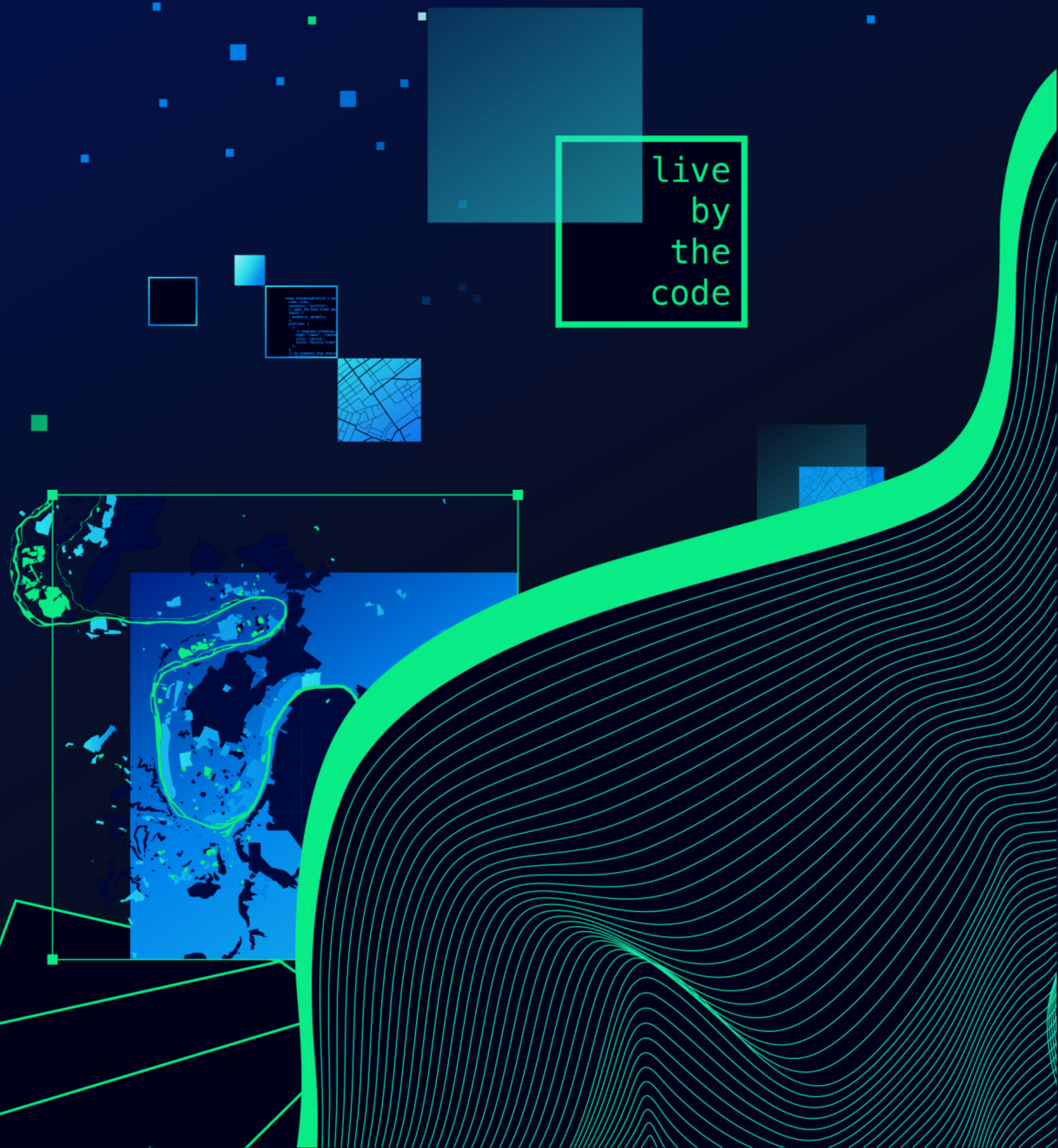
- The newer spatial vector representation in R
- Uses the simple features (sf) standard of OGC (Open Geospatial Consortium)
- Simple Features are:
    1. Point
    2. Polygon
    3. Linestring
    4. Multipoint
    5. Multipolygon
    6. Multilinestring
    7. Geometrycollection

Raster Data Representation in R

- Raster Info + Attributes
- Raster Info
  - nrows : Number of rows
  - ncols: Number of columns
  - nbands: Number of bands
  - extent
  - projection

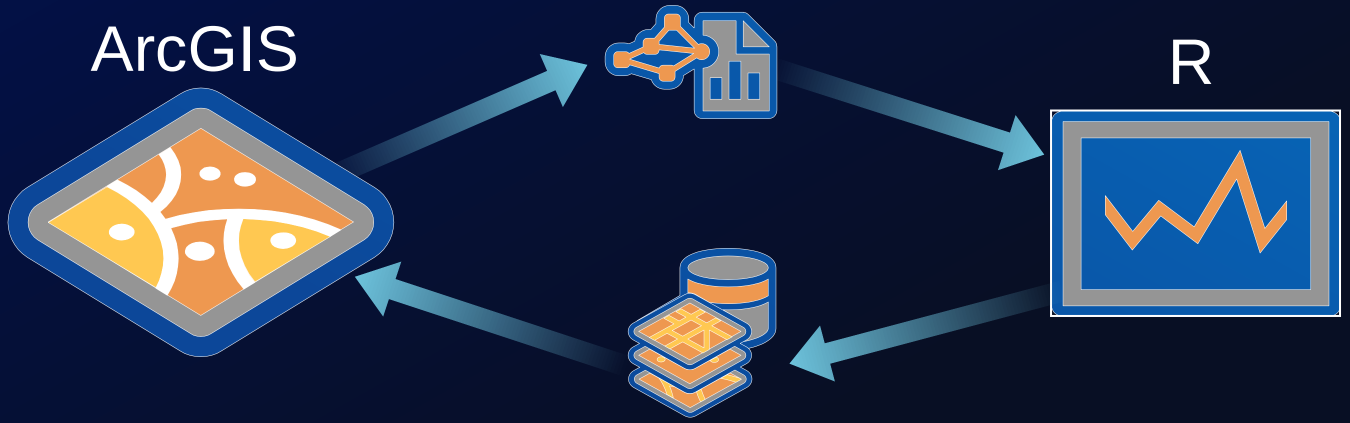
Connecting ArcGIS to R

The R-ArcGIS Bridge



R-ArcGIS Bridge

- Store your data in ArcGIS
- Directly access it in R
- Returns R objects back to ArcGIS native data types
- sf, sp, and raster conversion



Different Ways of Setting up the R-ArcGIS Bridge

Geoprocessing Tab

- Fastest way to get setup
- You need to have R installed
- Useful if you have R on your machine

Conda Package Manager

- Easy and streamlined process
- You do NOT need to have R installed
- Independent working environments

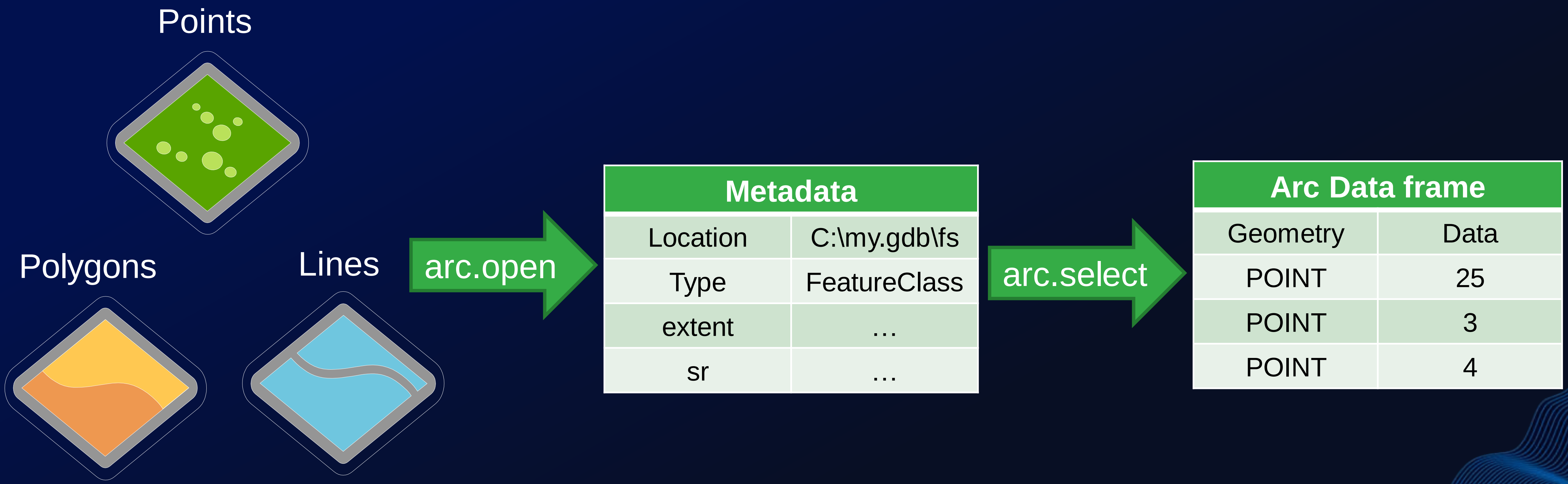
GitHub Repository

- Least streamlined
- Necessary for ArcGIS Desktop
- Try if everything else fails for ArcGIS Pro

arcgisbinding package

- Spatial data support
  - Vector Data (Shapefiles, File geodatabase, feature service)
  - Raster Data (Raster, Mosaic datasets)
- Read/Write Functionality
- Converting Functionality
- Wrapping R Tool Functionality

arcgisbinding functions for Vector Data



Metadata	
Location	C:\my.gdb\fs
Type	FeatureClass
extent	…
sr	…

Arc Data frame	
Geometry	Data
POINT	25
POINT	3
POINT	4

arc.open

- Function signature: ***arc.open***(path)
- **Input:** path, full path to data
    - Feature Class (ESRI)
    - Shapefile
    - Feature Service
    - OGC compliant data types
- **Output:** Metadata for data in path

arc.select

- Brings in data associated with metadata defined by arc.open
- ***arc.select***(object, fields, where_clause, selected, sr)
- **Input:**
  - object: Output from arc.open. Metadata
  - fields: List of fields to bring in
  - where_clause: SQL expression to bring in a subset of data
  - sr: Spatial reference for reprojecting
- **Output:**
  - Arc format spatial dataframe

☢ What is arc format?

arc.data2sf & arc.data2sp

- Converter functions

- Converts arc-type spatial R data frames into sf and sp type spatial dataframes
  - arc.data2sf(data_arc)
  - arc.data2sp(data_arc)

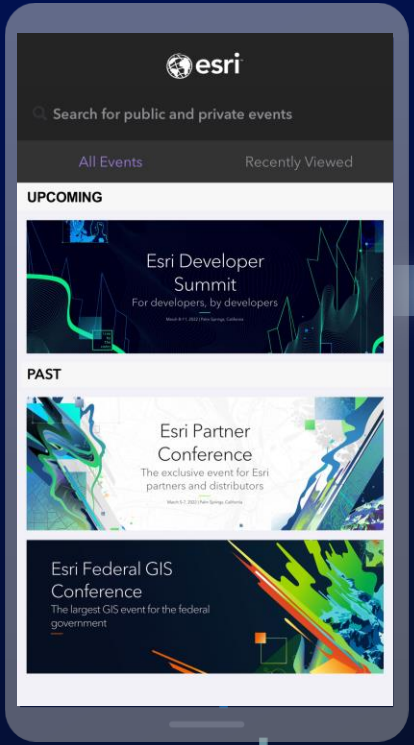
- **Input:** data_arc : arc type spatial R data frame
- **Output:** sp or sf type spatial R data frame

arc.write

- Writes spatial a spatial R data frame into disk
- Input data frame can be a:
  - sf object
  - sp object
  - arc object
- Output for vector data can be:
  - ESRI Feature Class
  - Shapefile

Please Share Your Feedback in the App

Download the Esri Events app and find your event



esri

Search for public and private events

All Events | Recently Viewed

UPCOMING

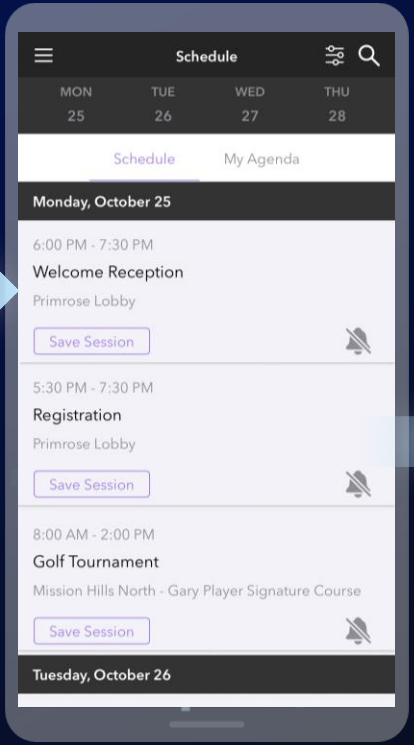
Esri Developer Summit
For developers, by developers

PAST

Esri Partner Conference
The exclusive event for Esri partners and distributors

Esri Federal GIS Conference
The largest GIS event for the federal government

Select the session you attended



Schedule

MON 25 | TUE 26 | WED 27 | THU 28

Schedule | My Agenda

Monday, October 25

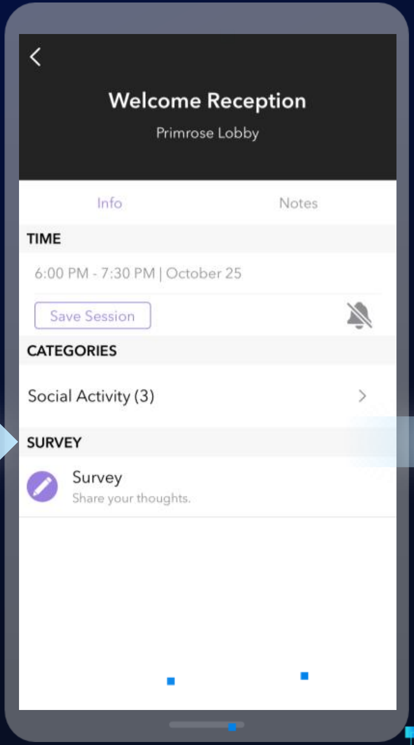
6:00 PM - 7:30 PM
Welcome Reception
Primrose Lobby
Save Session

5:30 PM - 7:30 PM
Registration
Primrose Lobby
Save Session

8:00 AM - 2:00 PM
Golf Tournament
Mission Hills North - Gary Player Signature Course
Save Session

Tuesday, October 26

Scroll down to “Survey”



Welcome Reception
Primrose Lobby

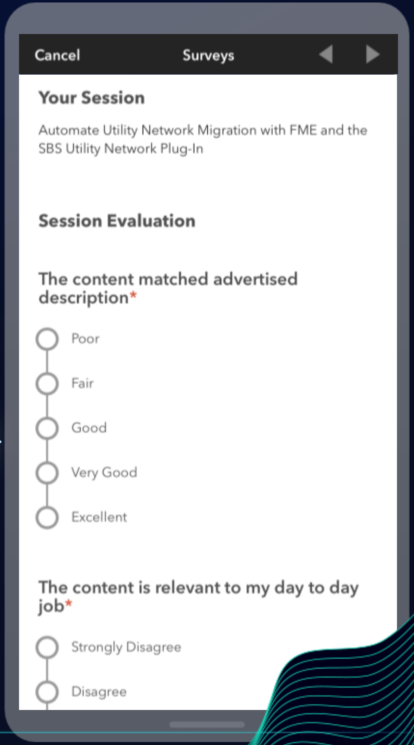
Info | Notes

TIME
6:00 PM - 7:30 PM | October 25
Save Session

CATEGORIES
Social Activity (3)

SURVEY
Survey
Share your thoughts.

Log in to access the survey



Cancel | Surveys

Your Session

Automate Utility Network Migration with FME and the SBS Utility Network Plug-In

Session Evaluation

The content matched advertised description*

○ Poor
○ Fair
○ Good
○ Very Good
○ Excellent

The content is relevant to my day to day job*

○ Strongly Disagree
○ Disagree

esri®

THE
SCIENCE
OF
WHERE®