R: The statistical programming language

Colin Smith
Earth Science Information Partners
Summer Meeting, July 17, 2019

R : The statistical programming language, that's a bit quirky and slow

Colin Smith
Earth Science Information Partners
Summer Meeting, July 17, 2019

R: The statistical programming language, that’s a bit quirky and slow, but extremely versatile

Colin Smith

Earth Science Information Partners
Summer Meeting, July 17, 2019

R : The statistical programming language, that’s a bit quirky and slow, but extremely versatile and has a lot of unrealized potential

Colin Smith

Earth Science Information Partners
Summer Meeting, July 17, 2019

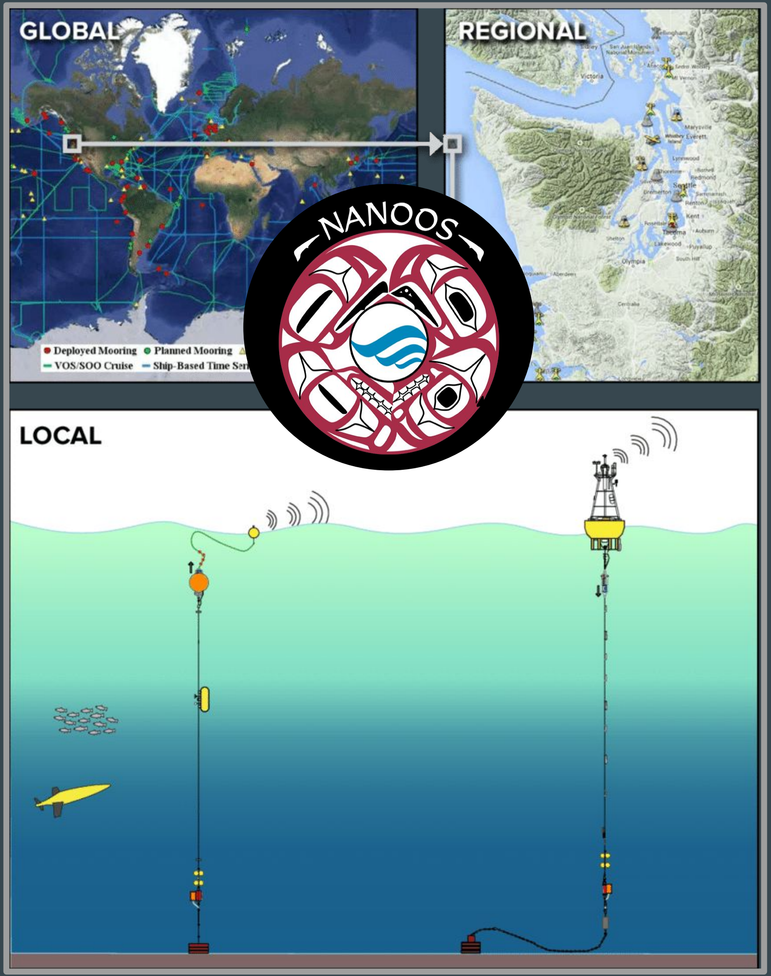
What I do

What I do

- Data curation

What I do

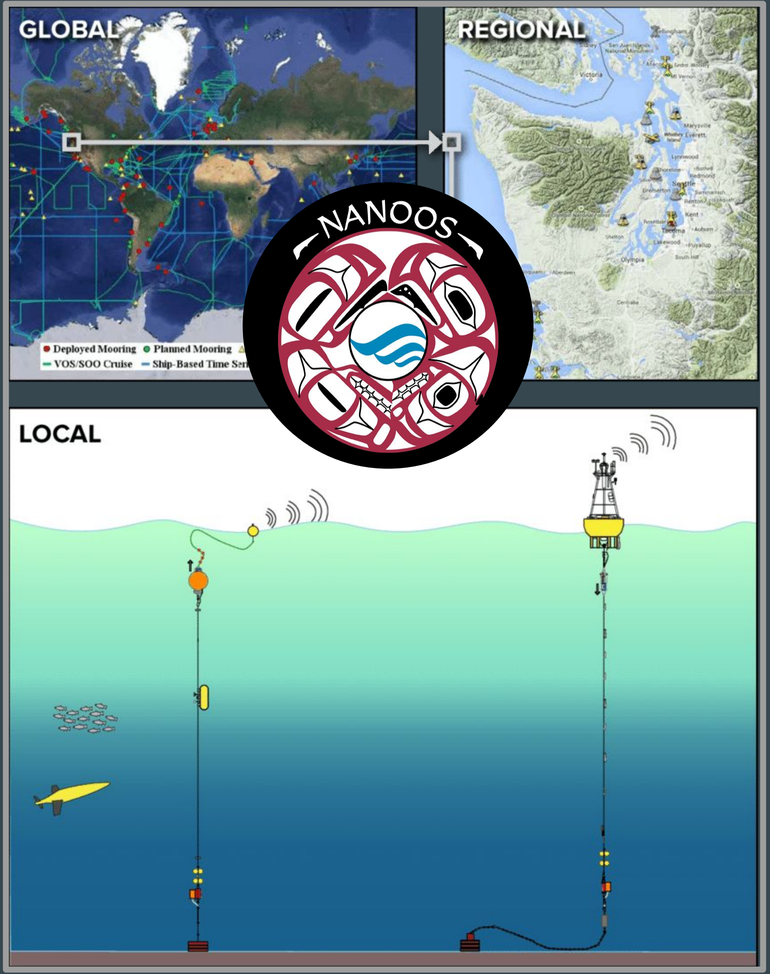
- Data curation
- Limnology and Oceanography



Alin et al. 2017. Characterizing the natural system: Toward sustained, integrated coastal ocean acidification observing networks to facilitate resource management and decision support.

What I do

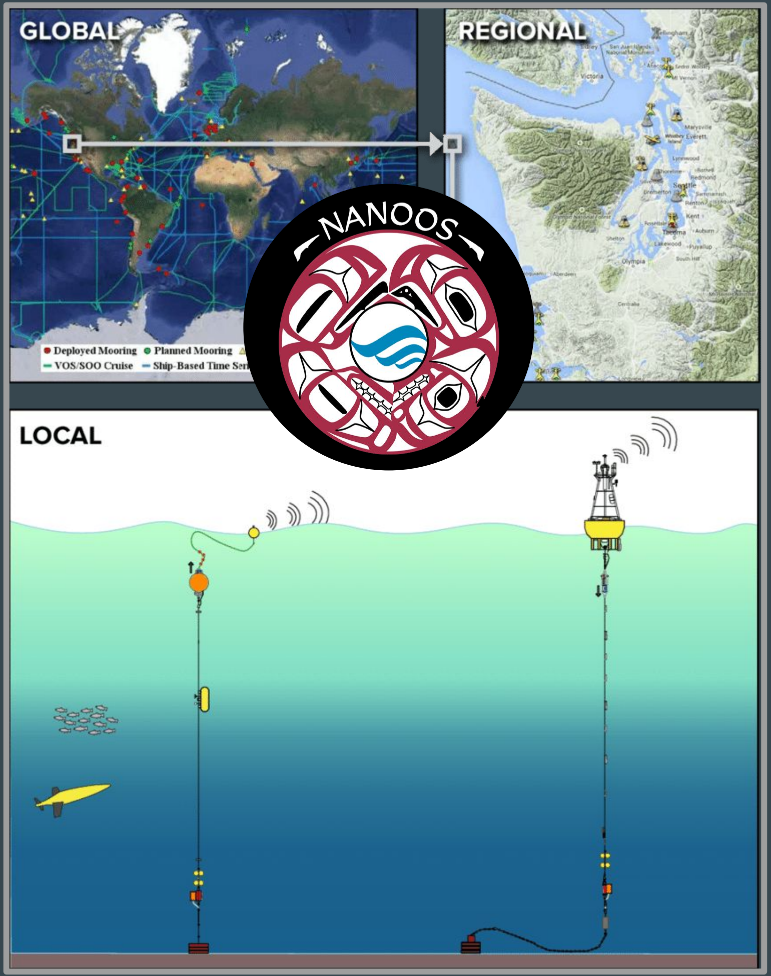
- Data curation
- Limnology and Oceanography
- Sensor networks



Alin et al. 2017. Characterizing the natural system: Toward sustained, integrated coastal ocean acidification observing networks to facilitate resource management and decision support.

What I do

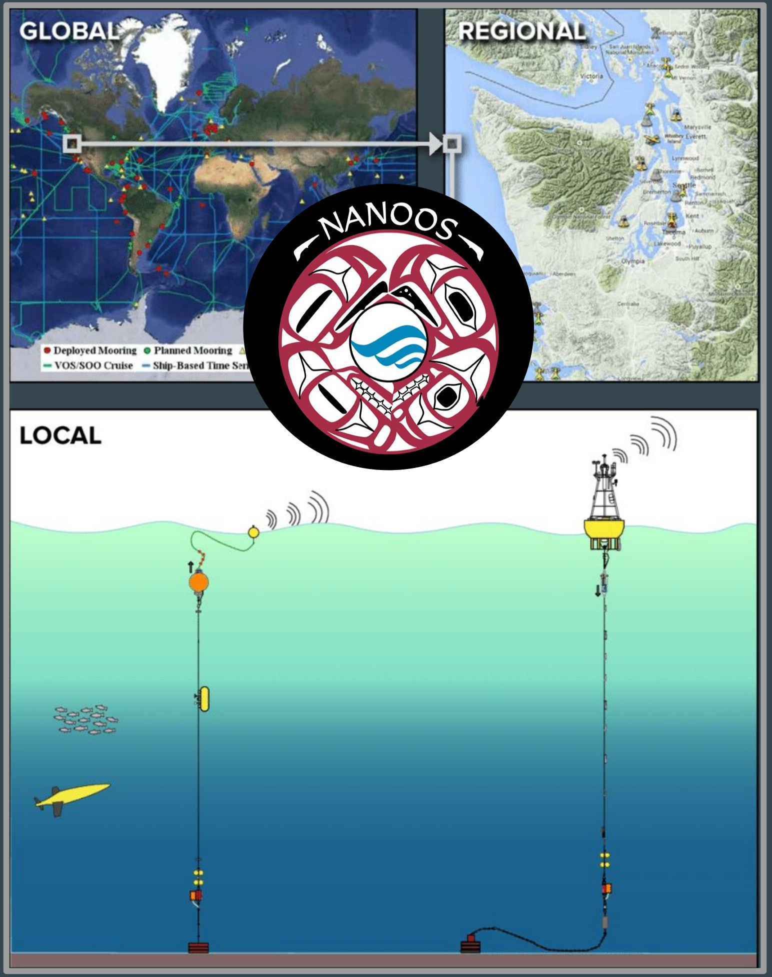
- Data curation
- Limnology and Oceanography
- Sensor networks
- Creative process



Alin et al. 2017. Characterizing the natural system: Toward sustained, integrated coastal ocean acidification observing networks to facilitate resource management and decision support.

What I do

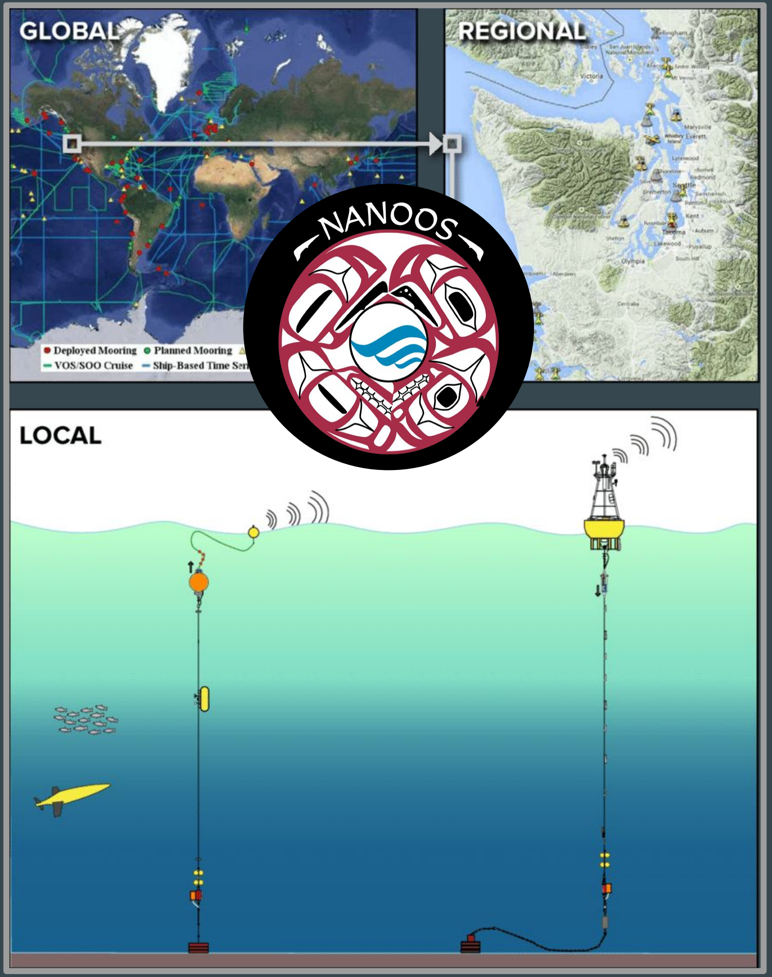
- Data curation
- Limnology and Oceanography
- Sensor networks
- Creative process
- Software development



Alin et al. 2017. Characterizing the natural system: Toward sustained, integrated coastal ocean acidification observing networks to facilitate resource management and decision support.

What I do

- Data curation
- Limnology and Oceanography
- Sensor networks
- Creative process
- Software development
- MATLAB, Perl, and R



Alin et al. 2017. Characterizing the natural system: Toward sustained, integrated coastal ocean acidification observing networks to facilitate resource management and decision support.

How I use R

How I use R

- Data processing

How I use R

- Data processing
- Visualization

How I use R

- Data processing
- Visualization
- Analysis

How I use R

- Data processing
- Visualization
- Analysis
- Data curation

How I use R

- Data processing
- Visualization
- Analysis
- Data curation
- File transfer

How I use R

- Data processing
- Visualization
- Analysis
- Data curation
- File transfer
- Web-services

How I use R

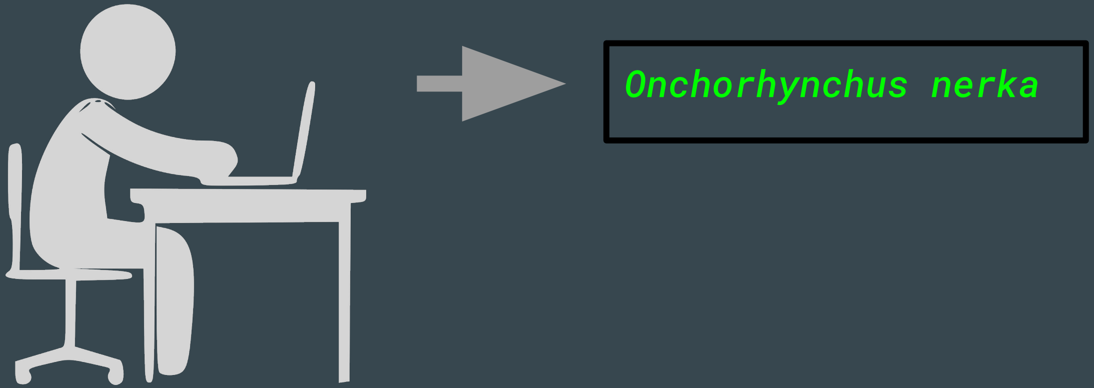
- Data processing
- Visualization
- Analysis
- Data curation
- File transfer
- Web-services
- Tooling

How I use R

- Data processing
- Visualization
- Analysis
- Data curation
- File transfer
- Web-services
- Tooling
- Workflows

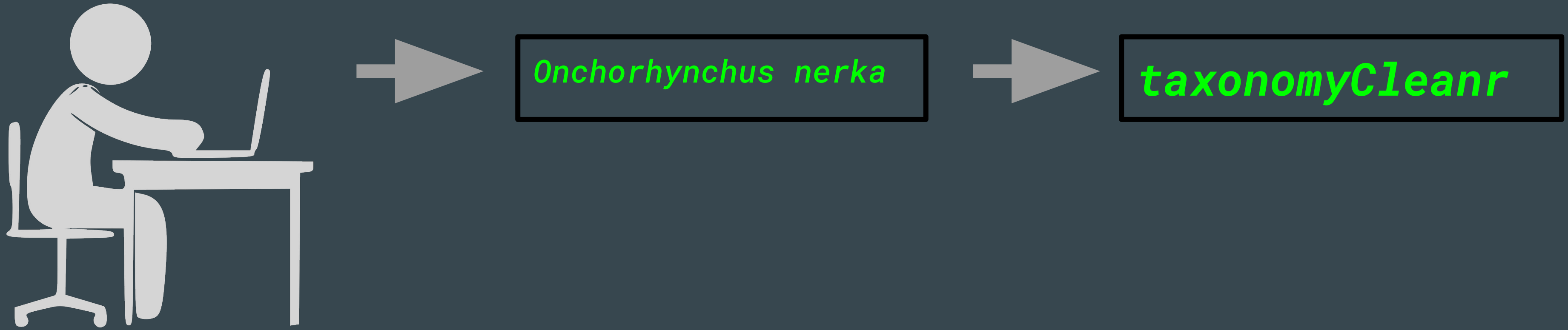
How I use R (taxonomyCleanr)

How I use R (taxonomyCleanr)



Onchorhynchus nerka

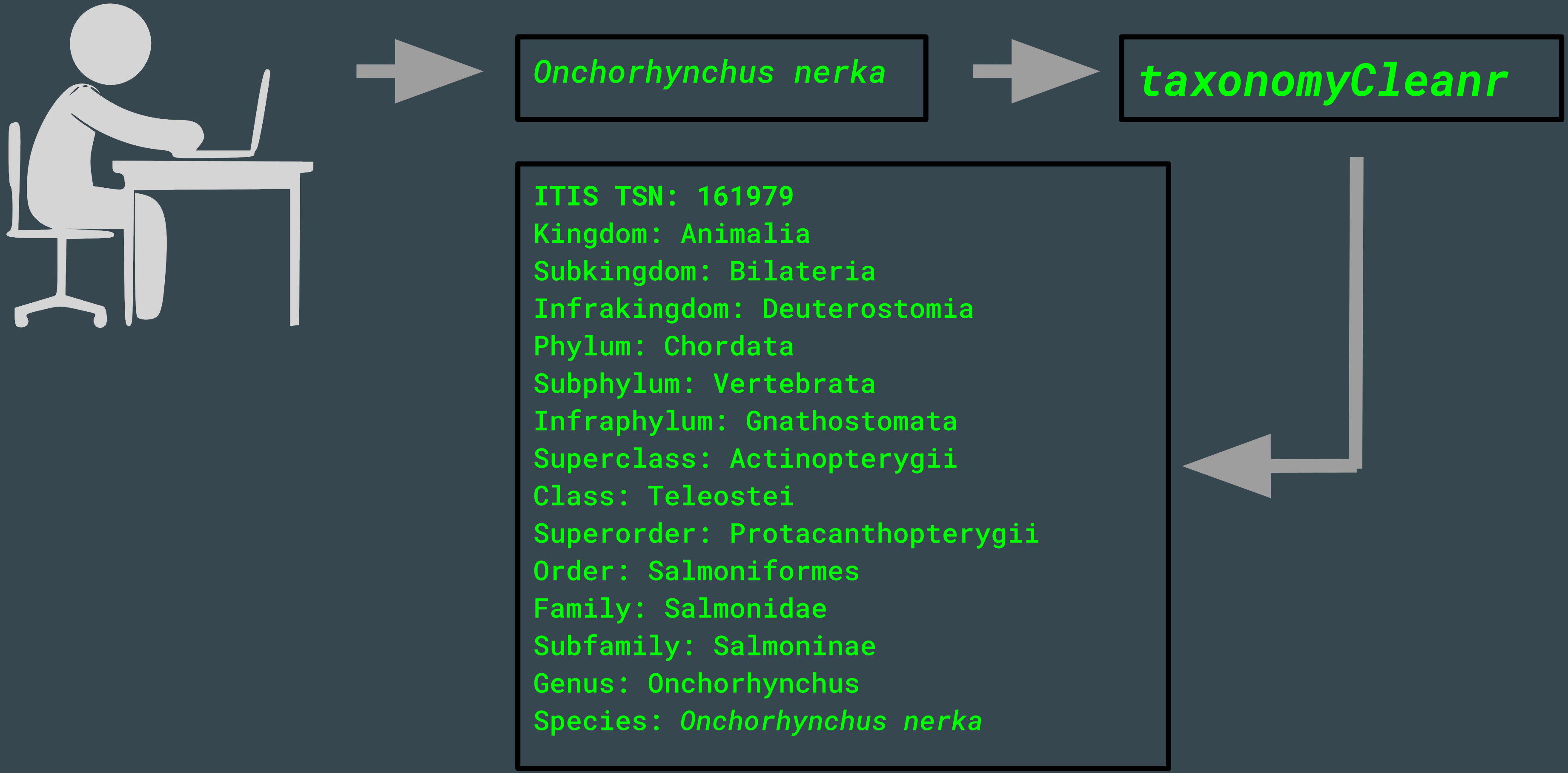
How I use R (taxonomyCleanr)



Onchorhynchus nerka

taxonomyCleanr

How I use R (taxonomyCleanr)

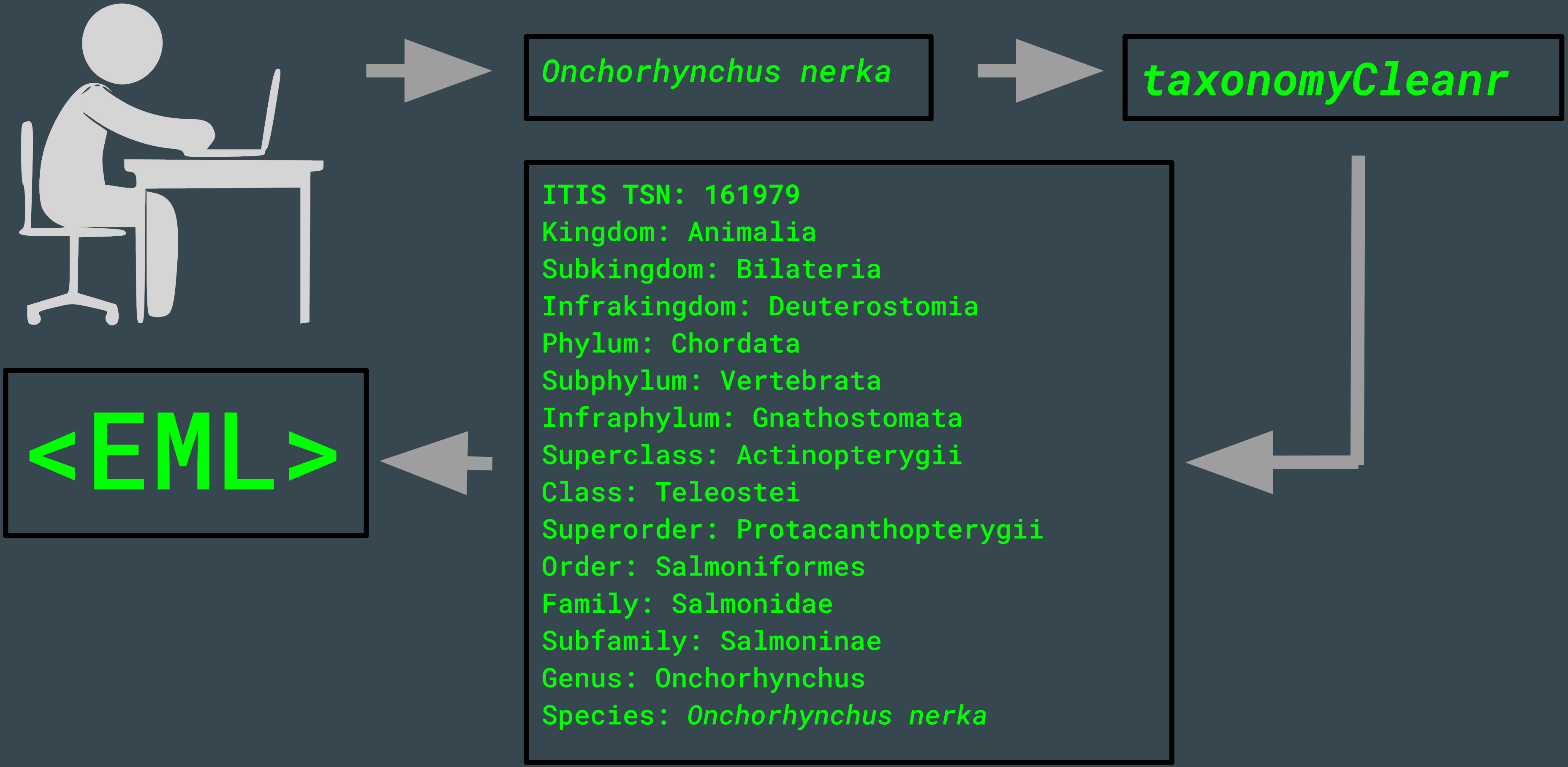


Onchorhynchus nerka

taxonomyCleanr

ITIS TSN: 161979
Kingdom: Animalia
Subkingdom: Bilateria
Infrakingdom: Deuterostomia
Phylum: Chordata
Subphylum: Vertebrata
Infraphylum: Gnathostomata
Superclass: Actinopterygii
Class: Teleostei
Superorder: Protacanthopterygii
Order: Salmoniformes
Family: Salmonidae
Subfamily: Salmoninae
Genus: Onchorhynchus
Species: *Onchorhynchus nerka*

How I use R (taxonomyCleanr)



Onchorhynchus nerka

taxonomyCleanr

ITIS TSN: 161979
Kingdom: Animalia
Subkingdom: Bilateria
Infrakingdom: Deuterostomia
Phylum: Chordata
Subphylum: Vertebrata
Infraphylum: Gnathostomata
Superclass: Actinopterygii
Class: Teleostei
Superorder: Protacanthopterygii
Order: Salmoniformes
Family: Salmonidae
Subfamily: Salmoninae
Genus: Onchorhynchus
Species: *Onchorhynchus nerka*

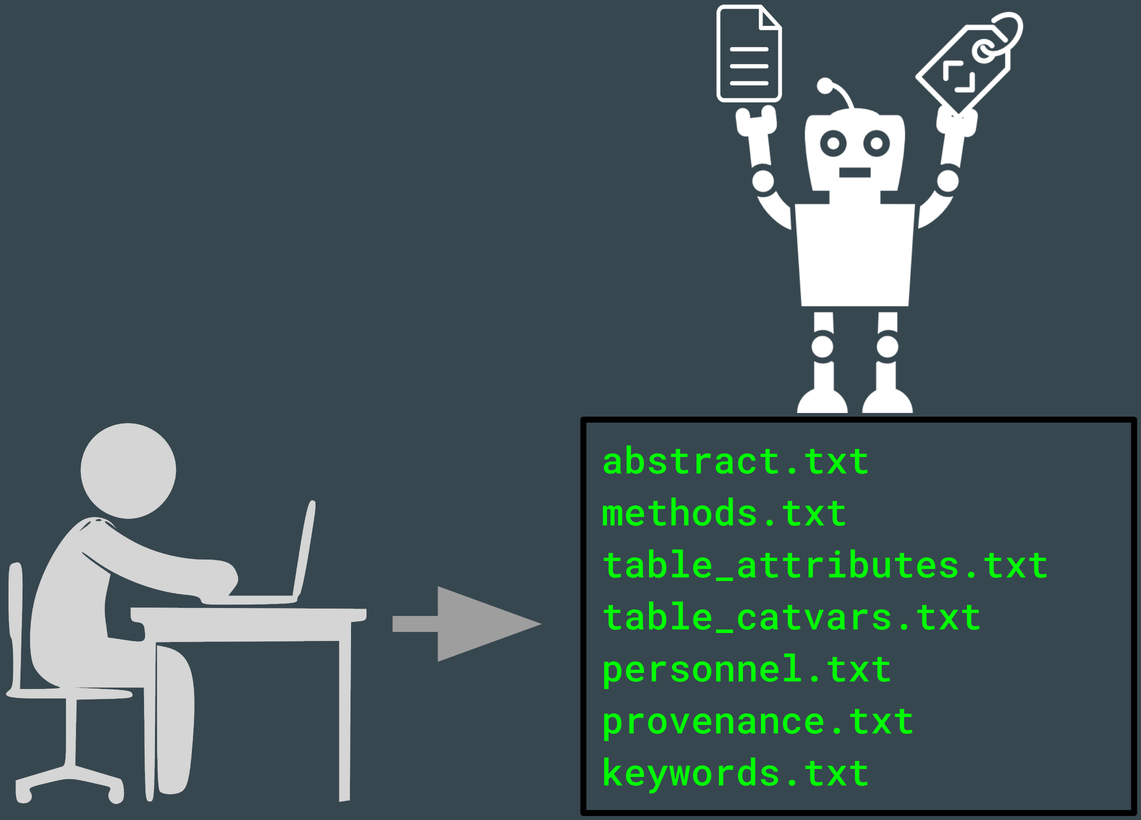
<EML>

How I use R (EMLassemblyline)

How I use R (EMLassemblyline)

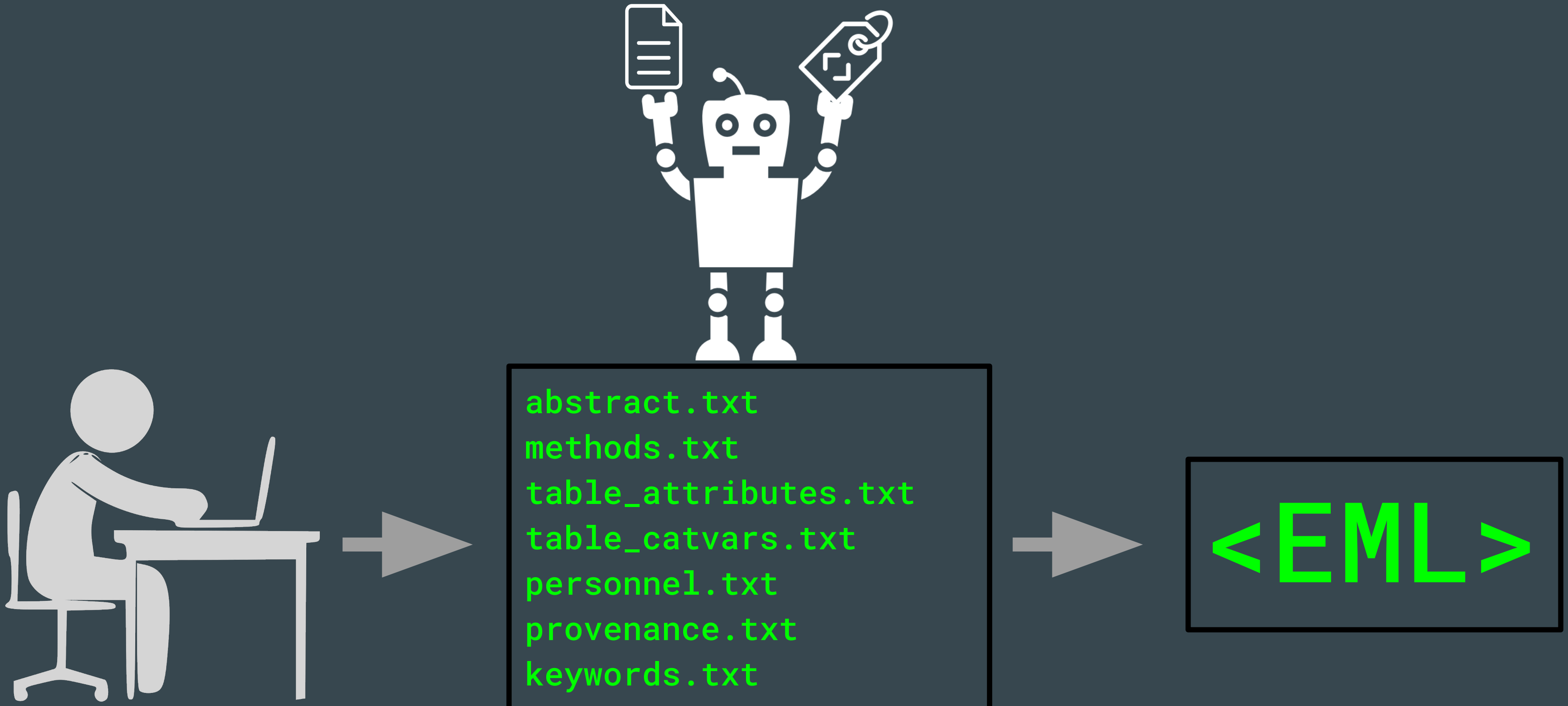
```
abstract.txt
methods.txt
table_attributes.txt
table_catvars.txt
personnel.txt
provenance.txt
keywords.txt
```

How I use R (EMLassemblyline)



```
abstract.txt
methods.txt
table_attributes.txt
table_catvars.txt
personnel.txt
provenance.txt
keywords.txt
```

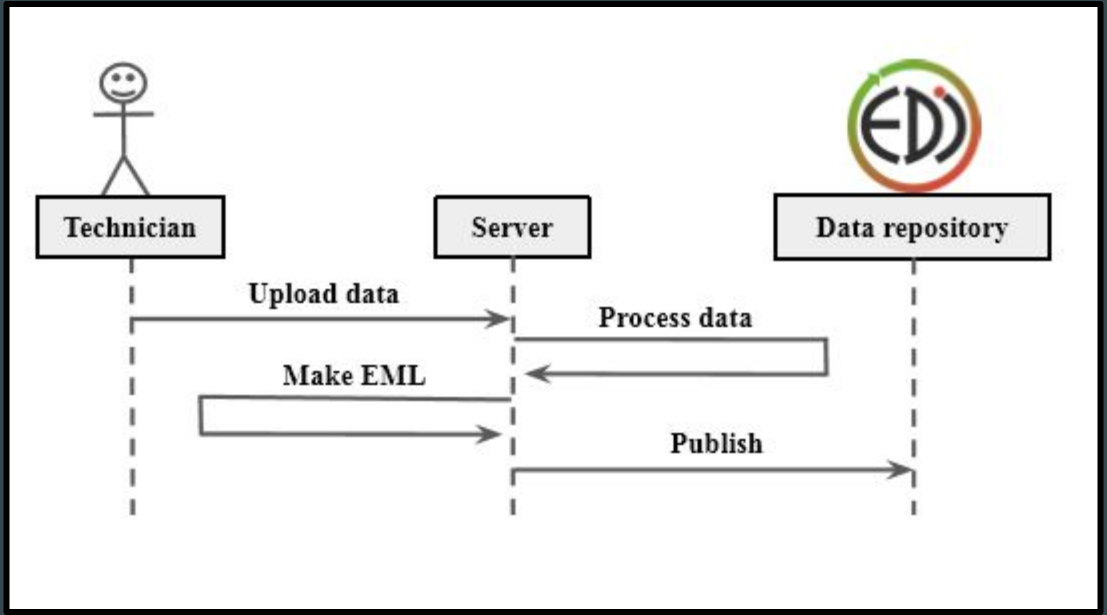
How I use R (EMLassemblyline)



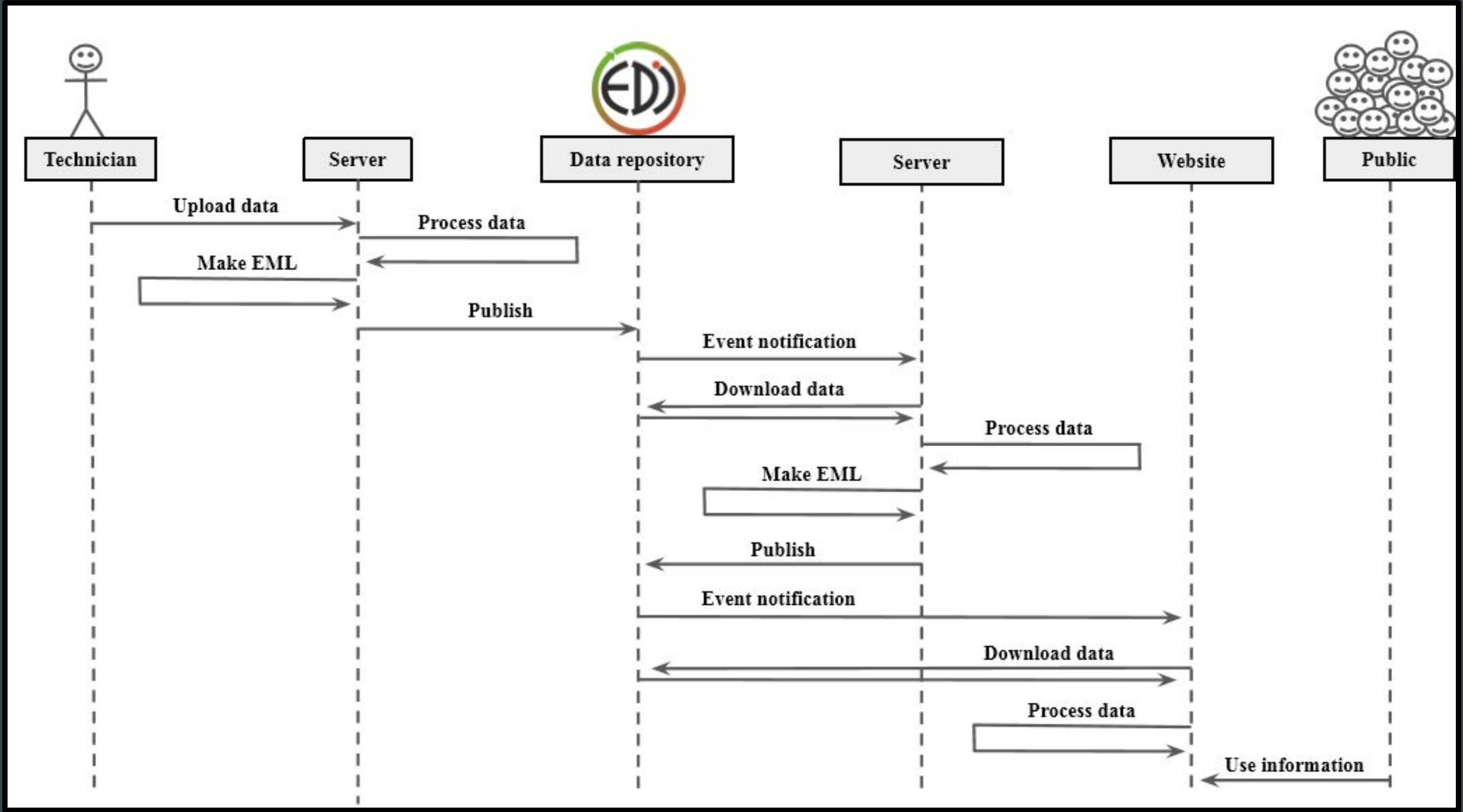
abstract.txt
methods.txt
table_attributes.txt
table_catvars.txt
personnel.txt
provenance.txt
keywords.txt

<EML>

How I use R (Automated data publishing)



How I use R (Reproducible workflows)



Major characteristics of R

Major characteristics of R

- Implementation of S

Major characteristics of R

- Implementation of S
- Interpreted language

Major characteristics of R

- Implementation of S
- Interpreted language
- Born August 1993

Major characteristics of R

- Implementation of S
- Interpreted language
- Born August 1993
- Ross Ihaka and Robert Gentleman

Major characteristics of R

- Implementation of S
- Interpreted language
- Born August 1993
- Ross Ihaka and Robert Gentleman
- Free and Open Source

Major characteristics of R

- Implementation of S
- Interpreted language
- Born August 1993
- Ross Ihaka and Robert Gentleman
- Free and Open Source
- Maintained by the R Core Team

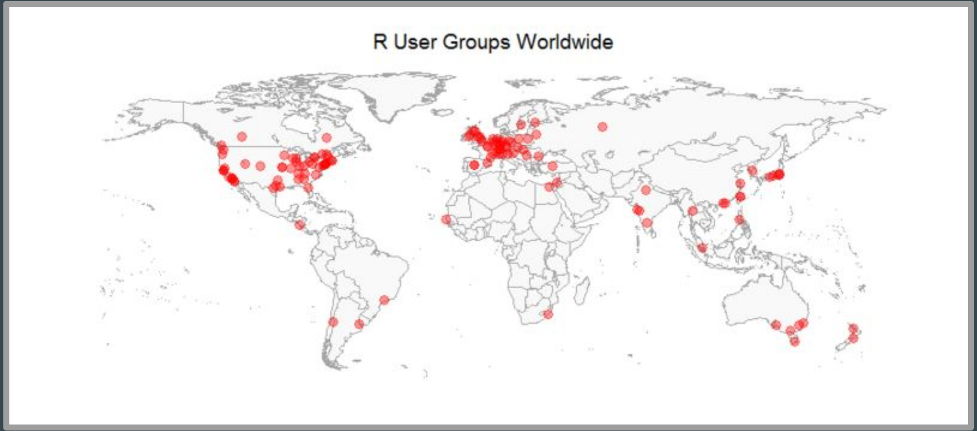
Major characteristics of R

- Implementation of S
- Interpreted language
- Born August 1993
- Ross Ihaka and Robert Gentleman
- Free and Open Source
- Maintained by the R Core Team
- Compatible with major OS

Major characteristics of R

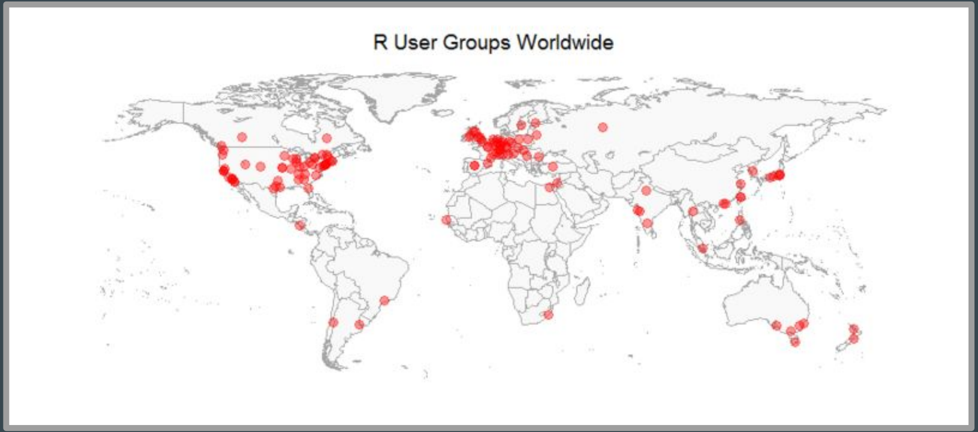
Major characteristics of R

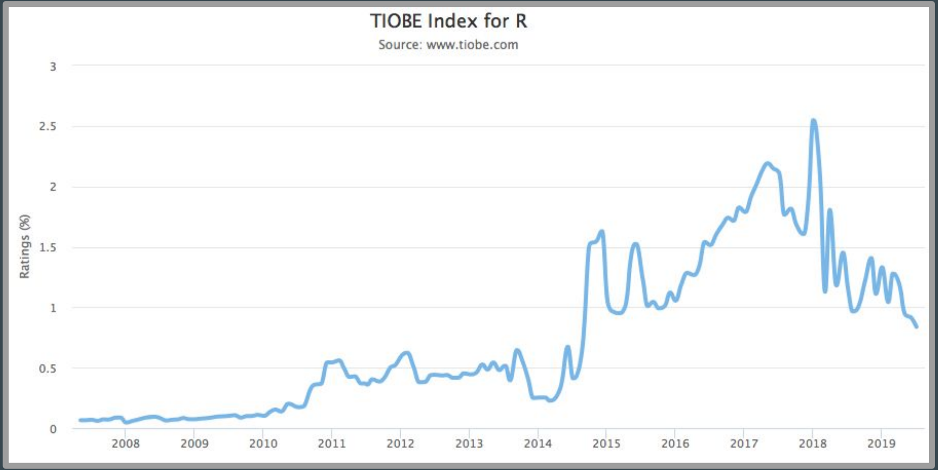
- Large user community



Major characteristics of R

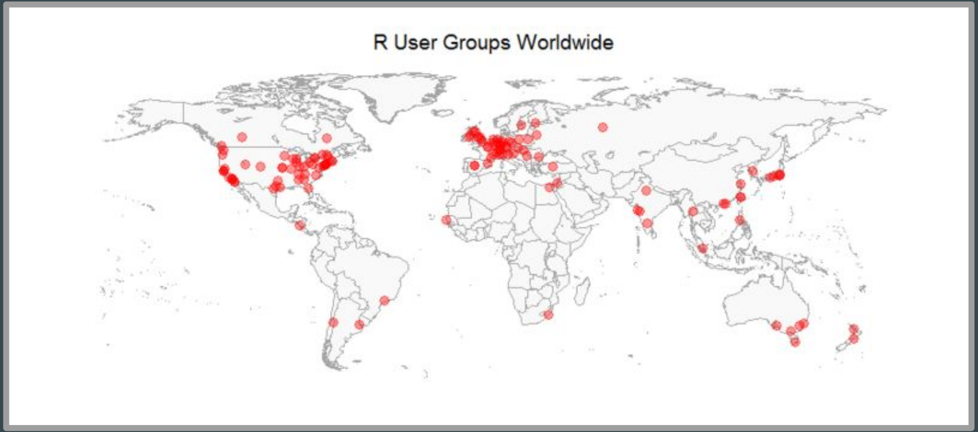
- Large user community

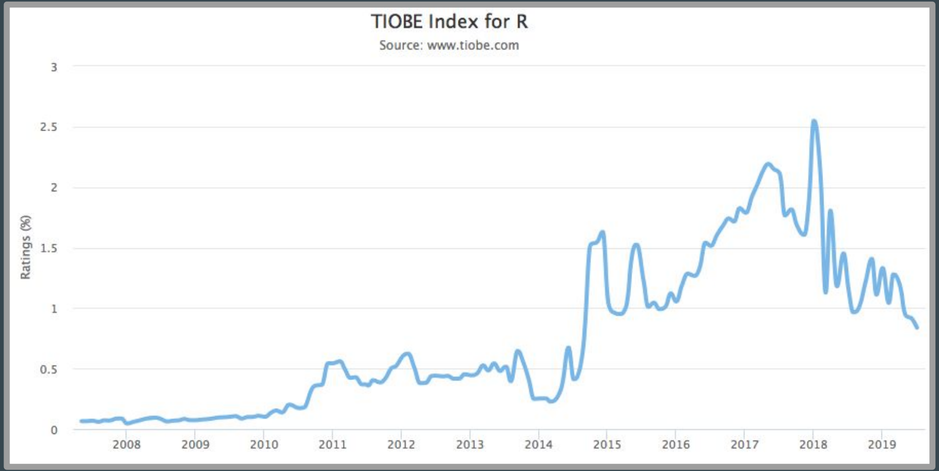




Major characteristics of R

- Large user community
- Academic

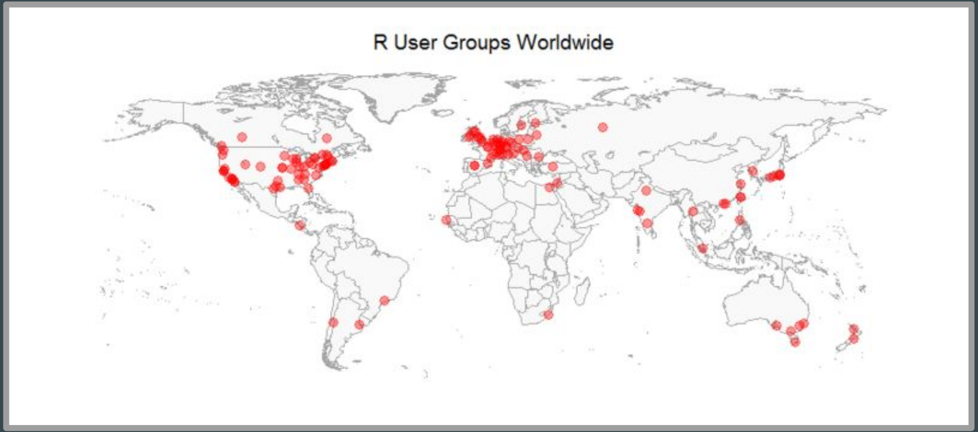

R User Groups Worldwide

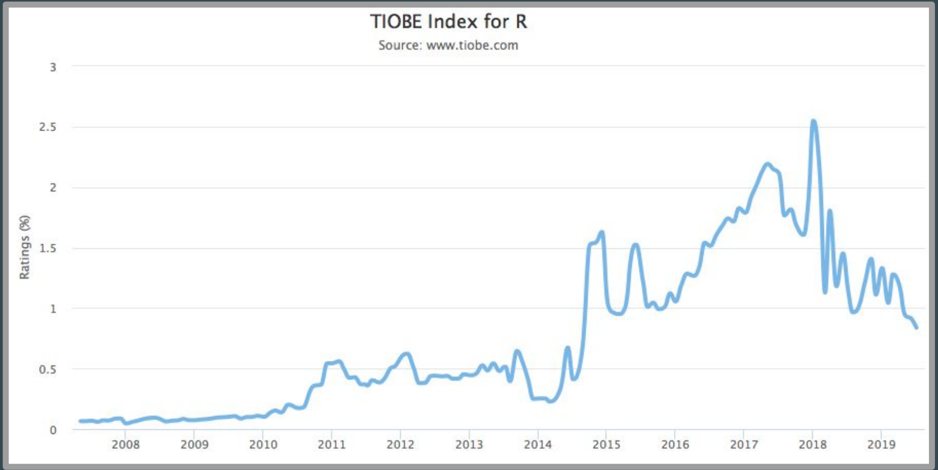

TIOBE Index for R

Source: www.tiobe.com

Major characteristics of R

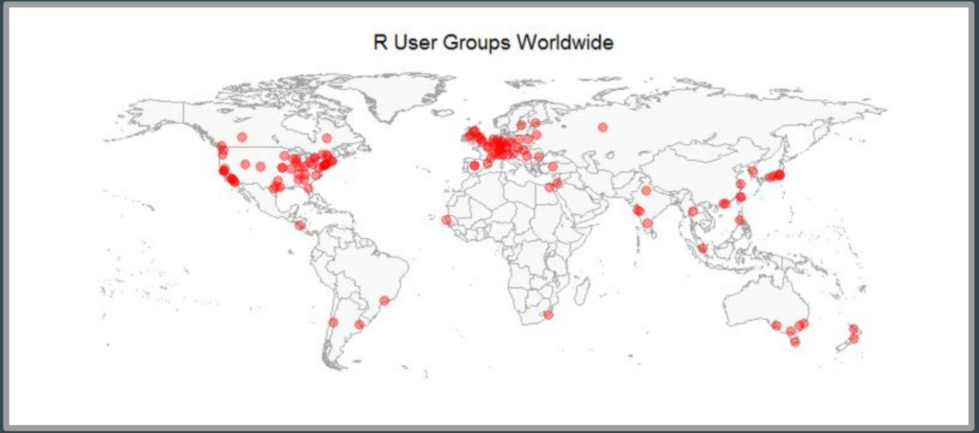
- Large user community
- Academic
- Commercial



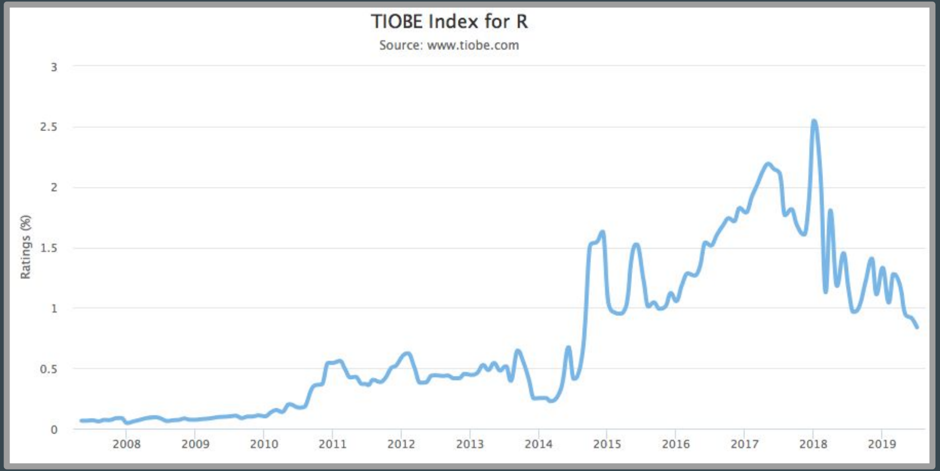


Major characteristics of R

- Large user community
- Academic
- Commercial
- R-Ladies, useR!, WhyR?



R User Groups Worldwide



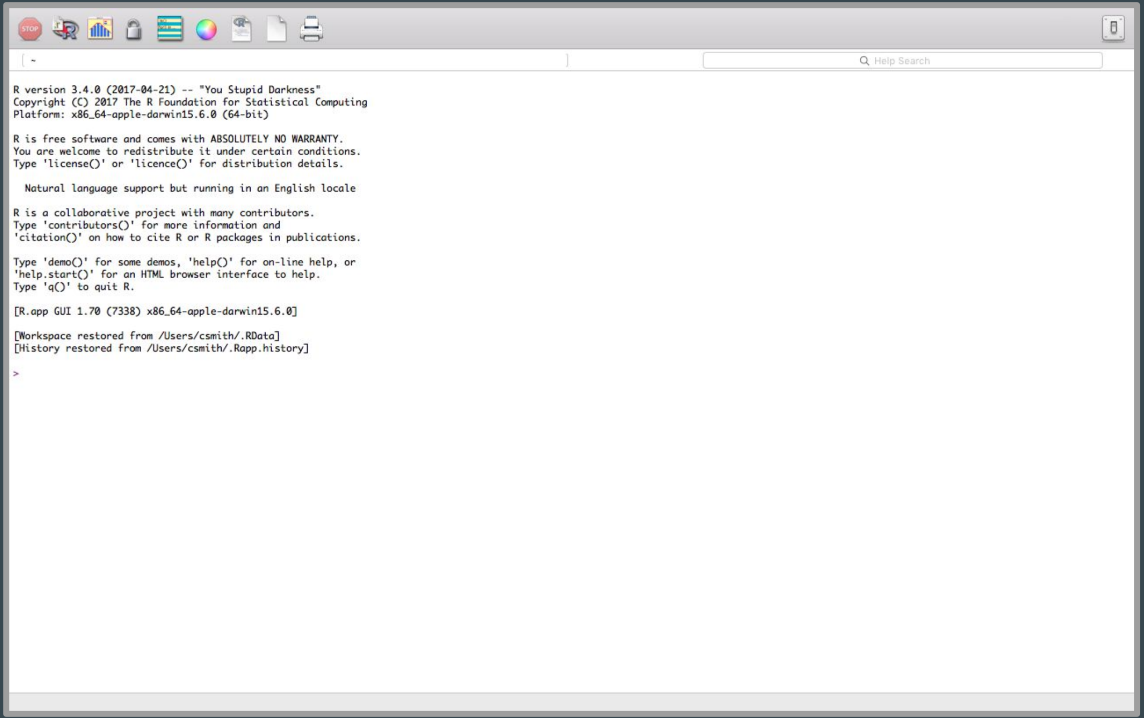
TIOBE Index for R

Source: www.tiobe.com

Major characteristics of R

Major characteristics of R

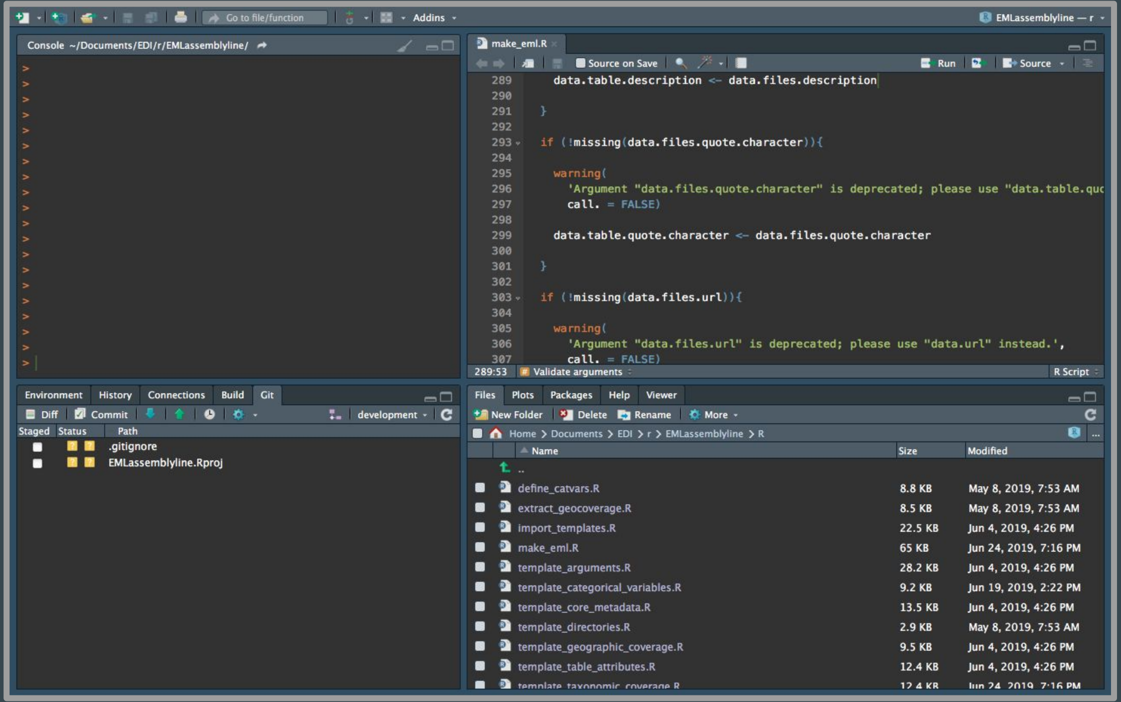
- Command line interface



```
R version 3.4.0 (2017-04-21) -- "You Stupid Darkness"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.70 (7338) x86_64-apple-darwin15.6.0]

[Workspace restored from /Users/csmith/.RData]
[History restored from /Users/csmith/.Rapp.history]

>
```

Major characteristics of R

- Command line interface
- RStudio IDE:
  - Scripting
  - Profiling
  - Debugging
  - Testing
  - Version control
  - Package development



Console ~/Documents/EDI/r/EMLassemblyline/

```
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
```

Staged	Status	Path
□	? ?	.gitignore
□	? ?	EMLassemblyline.Rproj

```r
    data.table.description <- data.files.description

  }

  if (!missing(data.files.quote.character)){

    warning(
      'Argument "data.files.quote.character" is deprecated; please use "data.table.quc
      call. = FALSE)

    data.table.quote.character <- data.files.quote.character

  }

  if (!missing(data.files.url)){

    warning(
      'Argument "data.files.url" is deprecated; please use "data.url" instead.',
      call. = FALSE)
```

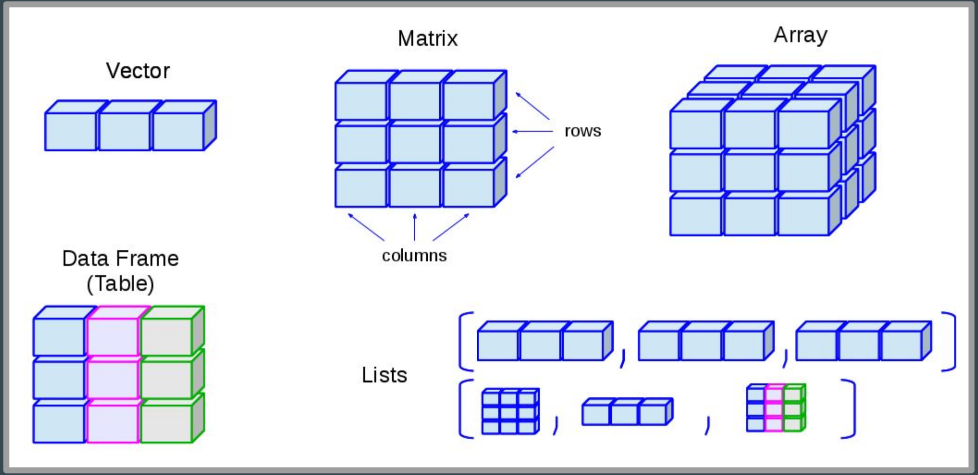
Home > Documents > EDI > r > EMLassemblyline > R

Name	Size	Modified
..		
define_catvars.R	8.8 KB	May 8, 2019, 7:53 AM
extract_geocoverage.R	8.5 KB	May 8, 2019, 7:53 AM
import_templates.R	22.5 KB	Jun 4, 2019, 4:26 PM
make_eml.R	65 KB	Jun 24, 2019, 7:16 PM
template_arguments.R	28.2 KB	Jun 4, 2019, 4:26 PM
template_categorical_variables.R	9.2 KB	Jun 19, 2019, 2:22 PM
template_core_metadata.R	13.5 KB	Jun 4, 2019, 4:26 PM
template_directories.R	2.9 KB	May 8, 2019, 7:53 AM
template_geographic_coverage.R	9.5 KB	Jun 4, 2019, 4:26 PM
template_table_attributes.R	12.4 KB	Jun 4, 2019, 4:26 PM
template_taxonomic_coverage.R	12.4 KB	Jun 24, 2019, 7:16 PM

Major characteristics of R

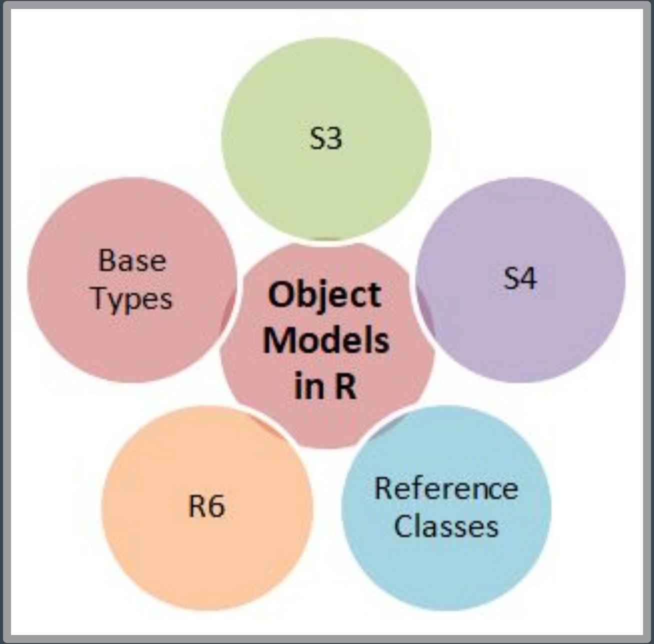
Major characteristics of R

- Data structures:
  - Vectors
  - Matrices
  - Arrays
  - Data frames
  - Lists
  - Environments



Major characteristics of R

- Data structures:
  - Vectors
  - Matrices
  - Arrays
  - Data frames
  - Lists
  - Environments
- Objects:
  - S3
  - S4
  - Reference classes
  - Base types
  - R6



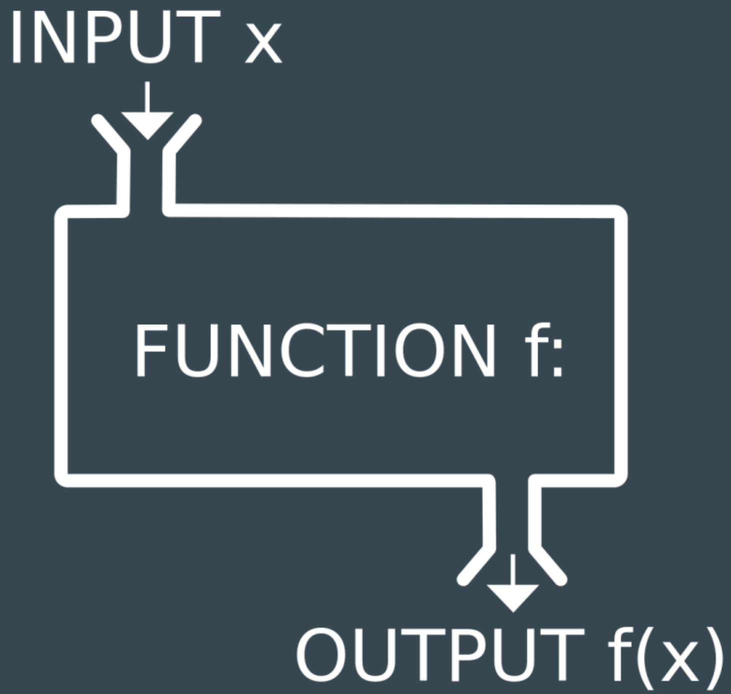
Major characteristics of R

Major characteristics of R

- Non-standard evaluation

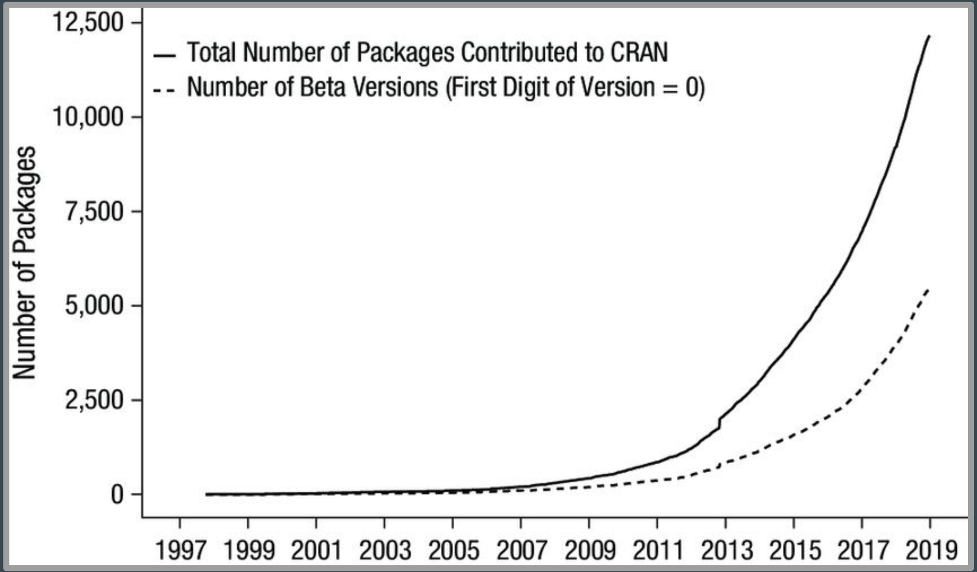
Major characteristics of R

- Non-standard evaluation
- Strong functional programming
  - Function are objects
  - Generic functions
  - Meta-programming
  - Dispatch
  - Lazy evaluation



Major characteristics of R

- Non-standard evaluation
- Strong functional programming
  - Function are objects
  - Generic functions
  - Meta-programming
  - Dispatch
  - Lazy evaluation
- Extensible
  - 15,000 packages
  - Easy to build and test



Epskamp. 2019. Reproducibility and Replicability in a Fast-Paced Methodological World.

Strengths of R

Strengths of R

- Available on major platforms

Strengths of R

- Available on major platforms
- Data-oriented problems

Strengths of R

- Available on major platforms
- Data-oriented problems
- Statistics

Strengths of R

- Available on major platforms
- Data-oriented problems
- Statistics
- Machine learning

Strengths of R

- Available on major platforms
- Data-oriented problems
- Statistics
- Machine learning
- Data manipulation

Strengths of R

- Available on major platforms
- Data-oriented problems
- Statistics
- Machine learning
- Data manipulation
- Visualization

Strengths of R

- Available on major platforms
- Data-oriented problems
- Statistics
- Machine learning
- Data manipulation
- Visualization
- Parallel computing

Strengths of R

- Available on major platforms
- Data-oriented problems
- Statistics
- Machine learning
- Data manipulation
- Visualization
- Parallel computing
- Reproducible research

Strengths of R

- Available on major platforms
- Data-oriented problems
- Statistics
- Machine learning
- Data manipulation
- Visualization
- Parallel computing
- Reproducible research
- Reports and websites

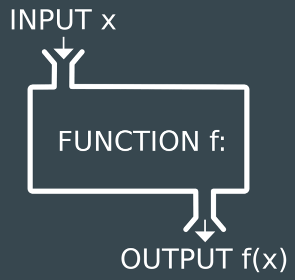
Strengths of R

- Available on major platforms
- Data-oriented problems
- Statistics
- Machine learning
- Data manipulation
- Visualization
- Parallel computing
- Reproducible research
- Reports and websites
- Versatile

Strengths of R

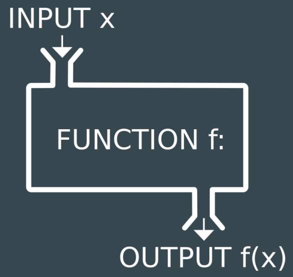
Strengths of R

- Functional programming



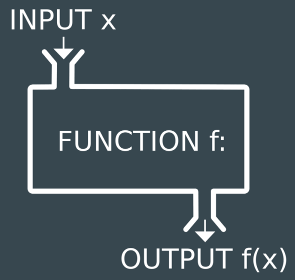
Strengths of R

- Functional programming
- Meta-programming



Strengths of R

- Functional programming
- Meta-programming
- Connects to other languages



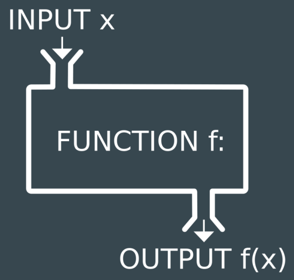
INPUT x

FUNCTION f:

OUTPUT f(x)

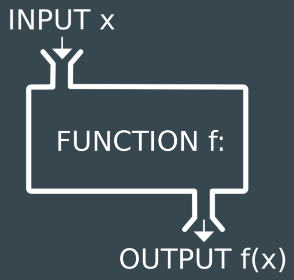
Strengths of R

- Functional programming
- Meta-programming
- Connects to other languages
- Large user community



Strengths of R

- Functional programming
- Meta-programming
- Connects to other languages
- Large user community
- Cutting edge academic research



INPUT x

FUNCTION f:

OUTPUT f(x)

Weaknesses of R

Weaknesses of R

- Most users are not programmers

Weaknesses of R

- Most users are not programmers
- Readability of meta programming

Weaknesses of R

- Most users are not programmers
- Readability of meta programming
- Core package inconsistencies

Weaknesses of R

- Most users are not programmers
- Readability of meta programming
- Core package inconsistencies
- Framework integration

Weaknesses of R

- Most users are not programmers
- Readability of meta programming
- Core package inconsistencies
- Framework integration
- Security

Weaknesses of R

- Most users are not programmers
- Readability of meta programming
- Core package inconsistencies
- Framework integration
- Security
- Slower than alternatives

Weaknesses of R

- Most users are not programmers
- Readability of meta programming
- Core package inconsistencies
- Framework integration
- Security
- Slower than alternatives
- Memory management

Realizing R’s full potential

Realizing R’s full potential

- R Core Team

Realizing R’s full potential

- R Core Team
- Ecosystems (tidyverse)

Realizing R's full potential

- R Core Team
- Ecosystems (tidyverse)
- Implementations:
  - pqR (pretty quick R)
  - Renjin
  - FastR
  - Riposte

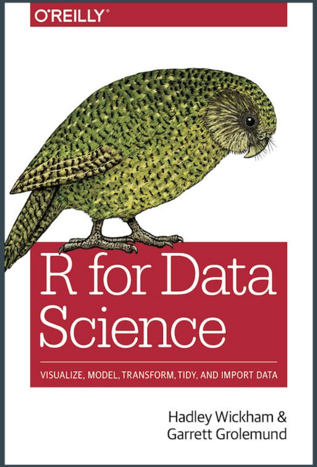
oracle / fastr

jtalbot / riposte

Resources

Resources

- R for data science


O'REILLY

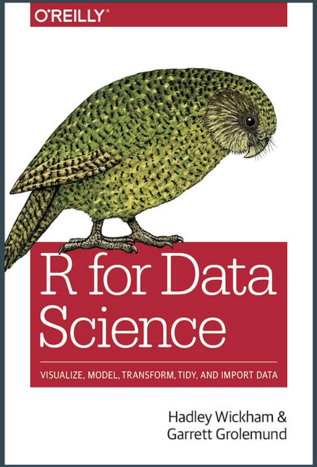
R for Data Science

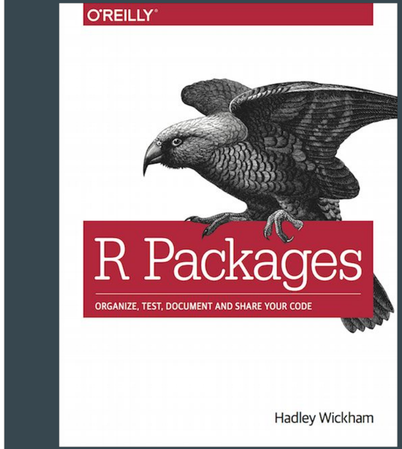
VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham & Garrett Grolemund

Resources

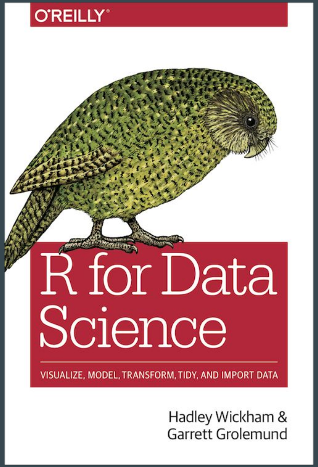
- R for data science
- R packages

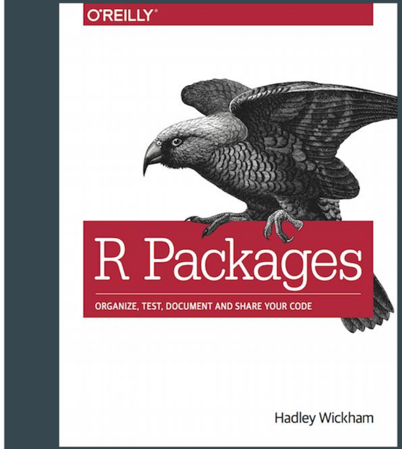


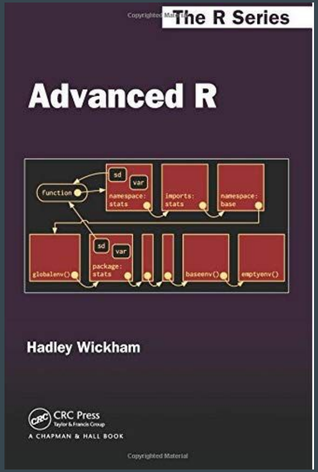


Resources

- R for data science
- R packages
- Advanced R

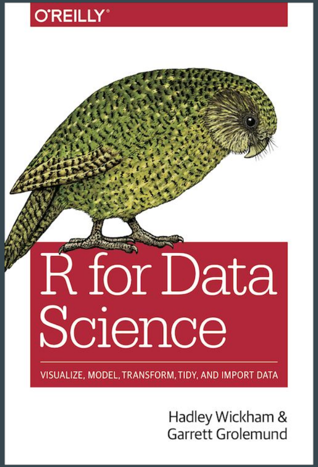


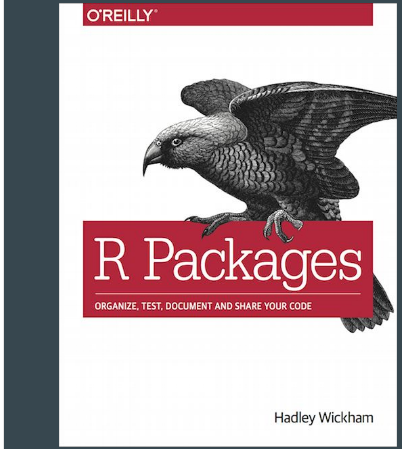


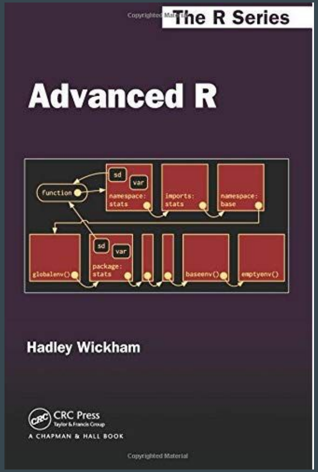


Resources

- R for data science
- R packages
- Advanced R
- R journal







The R Journal

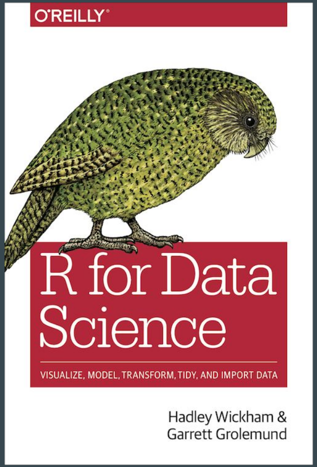
Volume 3/2, December 2011

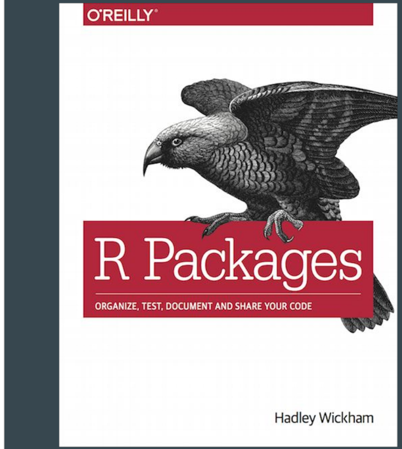
A peer-reviewed, open-access publication of the R Foundation for Statistical Computing

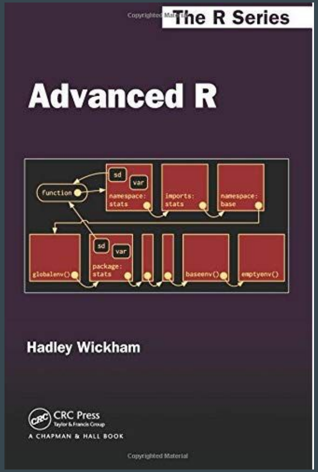
Contents

Resources

- R for data science
- R packages
- Advanced R
- R journal
- Web







The R Journal

Volume 3/2, December 2011

A peer-reviewed, open-access publication of the R Foundation
for Statistical Computing

Contents

“R”