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

How have you used the language?
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)
How I use R (taxonomyCleanr)

- **Onchorhynchus nerka**
- **taxonomyCleanr**

**ITIS TSN:** 161979
- **Kingdom:** Animalia
- **Subkingdom:** Bilateria
- **Infrakingdom:** Deuterostomia
- **Phylum:** Chordata
- **Subphylum:** Vertebrata
- **Infra phylum:** Gnathostomata
- **Superc lass:** Actinopterygii
- **Class:** Teleostei
- **Superorder:** Protacanthopterygii
- **Order:** Salmoniformes
- **Family:** Salmonidae
- **Subfamily:** Salmoninae
- **Genus:** Onchorhynchus
- **Species:** Onchorhynchus nerka
How I use R (taxonomyCleanr)

Onchorhynchus nerka

taxonomyCleanr

<EMML>

ITIS TSN: 161979
Kingdom: Animalia
Subkingdom: Bilateria
Infrakingdom: Deuterostomia
Phylum: Chordata
Subphylum: Vertebrata
Infra phylum: Gnathostomata
Superclass: Actinopterygii
Class: Teleostei
Superorder: Protacanthopterygii
Order: Salmoniformes
Family: Salmonidae
Subfamily: Salmoninae
Genus: Onchorhynchus
Species: Onchorhynchus nerka
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
What are the major characteristics of the language (R)?

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
Major characteristics of R

- Large user community
- Academic
- Commercial
What are the major characteristics of the language (R)?

- Large user community
- Academic
- Commercial
- R-Ladies, useR!, WhyR?
What are the major characteristics of the language (R)?
Major characteristics of R

- Command line interface
Major characteristics of R

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

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
What are the strengths of the language? (continued)

- 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
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
What are the weaknesses of the language?

- 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
Resources
Resources

- R for data science
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
Resources

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