



Introduction To R

EPA's Emerging Leaders Network
2021-02-24

Jeff Hollister,
ORD/CEMM/ACESD

Who am I?

- Landscape ecology, GIS, Lakes, and Data Science
- US EPA Research Ecologist since ~2006
- Worked on: Forests, wetlands, gopher tortoise, estuaries, lakes, cyanobacteria, ...
- useR since 2000 (yikes!)
- @jhollist, #rstats



What is R?

- “Free software environment for statistical computing and graphics”
- Ross Ihaka and Robert Gentleman
- Dialect of the S language
- R’s first release: 1993
- Version 1.0: 2000
- Current version 4.0.4, Feb 2021.
 - aka “Lost Library Book”
- Excels at statistics and visualization
- Fully functional, general purpose programming language
- Why is it called “R”?



Why use R?

- Free!
- Statistics and Graphics
- GIS/Spatial
- Writing Papers
- Presentations (like this one)
- Programming
- Develop new tools
- Reproducibility
- Open Science



michael barber
@mbarber83

I want one but it costs \$5,000 per year.



Stata ✓ @Stata · 22h

Anyone else getting creative with their face masks? Post yours below!



R vs. Python

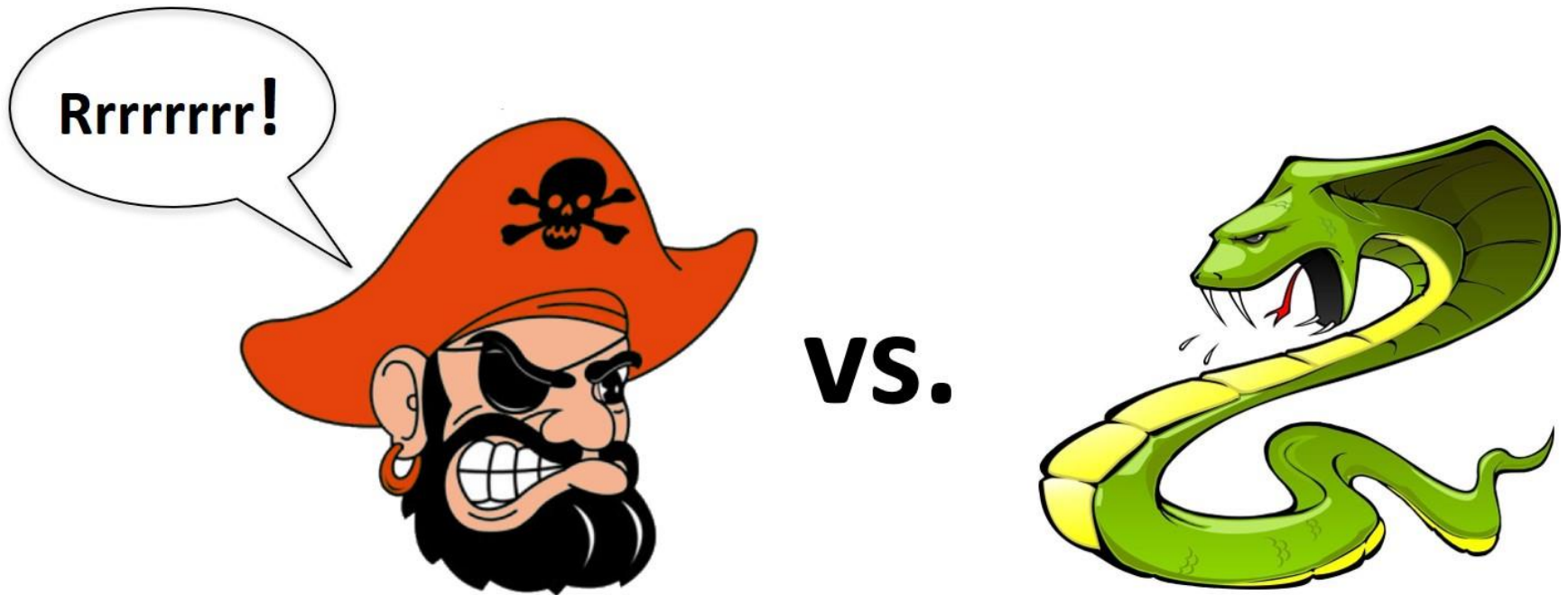


Image credits: <https://library.kissclipart.com/> and <https://pixy.org/896991/>

R vs. & Python



Image credit: https://github.com/rstudio/reticulate/blob/master/images/reticulated_python.png

- RStudio Panel discussion from yesterday: <https://www.youtube.com/watch?v=6mp-BWPnqK4>

R at EPA

- Large and active community
- July 2020: 1142 installs
 - HQ offices only
- [R Users Group](#)
 - Started ~2013/2014
 - Channel in Data Science team
 - Sharepoint (legacy)
 - Monthly meetings
 - Bi-annual face to face
- Installation (for HQ Offices, regions are Ad hoc)
 - R
 - RStudio
 - Rtools (required build tools for Windows)



One Minute Demos

- Integrated Development Environment
- Also a company that supports the R community
- RStudio 1-minute Demo

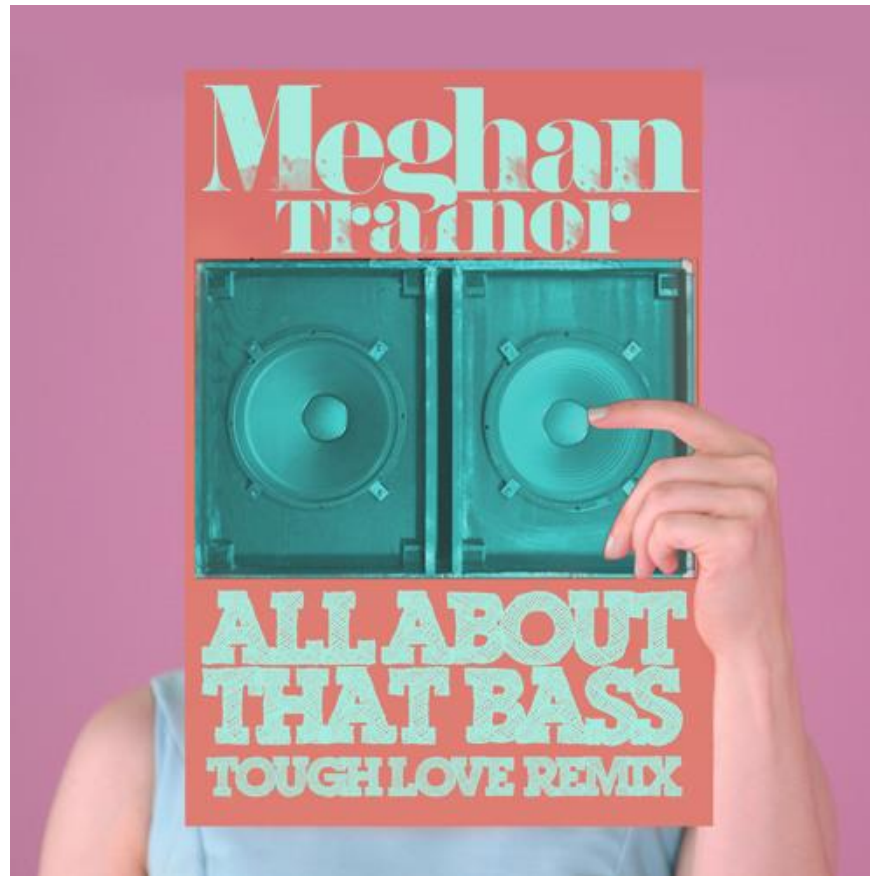
The screenshot displays the RStudio Integrated Development Environment (IDE) interface. The main editor window shows an R script for a function named `groan`, which is designed to return random dad jokes. The script includes comments, a URL for fetching jokes, and a `request` function to fetch jokes from a specific API endpoint. The console window at the bottom shows the execution of the `dadjokeapi::groan_search("fish")` function, which returns a tibble with 10 rows and 2 columns: `joke` and `id`. The file explorer on the right shows the project structure, including a `cran_comments.md` file and a `tests` directory.

```
1 # Function to return random dad joke
2 #
3 # This function returns a random dad joke(s) from
4 # \url{https://icanhazdadjoke.com}. as a message and invisibly as a character.
5 #
6 # @param sting Plays a joke sting after the joke. Sound from
7 # \url{https://archive.org/details/Rimshot_254}
8 # @export
9 # @return Invisibly returns a two item list with the first item holding the
10 # joke as a character and the second item containing the joke id as a
11 # character.
12 # @examples
13 # groan(sting = FALSE)
14 groan <- function(sting = TRUE) {
15
16   if (!curl::has_internet()) {
17     stop("Why did the chicken cross the road? Because you don't currently have an
18     internet connection.")
19   }
20
21   request <- http::GET("https://icanhazdadjoke.com")
22 }
```

```
> dadjokeapi::groan_search("fish")
# A tibble: 10 x 2
  joke id
<chr> <chr>
1 "why do fish live in salt water? Because pepper makes them sneeze!" q4hiGJBXLe
2 "what's the difference between a guitar and a fish? You can tune a gui- 8prwvkOf2~
3 "Two fish are in a tank, one turns to the other and says, \"how do you- aFtzPRSNb~
4 "what do you call a fish with no eyes? A fsh." 9EBljQWDA~
5 "why are fish easy to weigh? Because they have their own scales." KBsrz5ws4~
6 "what do you call two barracuda fish? A Pairacuda!" 1gyI6EIRK~
7 "why are fish so smart? Because they live in schools!" 1ge2Tvskyd
8 "what did the fish say when it swam into a wall? Damn!" axPR79MZvc
9 "Two parrots are sitting on a perch. One turns to the other and asks, ~ BAscNus4o~
10 "why do crabs never give to charity? Because they're shellfish." fqWgFt4Ed~
```

Base R

- Fresh out of the box you get 6618 functions!
- You can do a lot
 - stats, graphics, parallel computing, ...
- Base R 1-minute Demo



Extending Base R: Packages

- Community developed extensions
- Many quality checks
- Two main sources
 - CRAN (<https://cran.r-project.org/>): 17147 packages
 - Bioconductor (<https://www.bioconductor.org/>): 20463 packages
 - Genomics/bioinformatics



What can you do with R?

- Well-known:
 - Statistics
 - Graphics
- Less well-known:
 - Geospatial/GIS
 - Documents
 - word, pdf, ppt
 - Web applications
- Offbeat things



Image credit: <https://www.flickr.com/people/100030176@N02>

Offbeat 1 minute demo

- [beepR](#): Easily Play Notification Sounds on any Platform
- [BRRR](#): Rap adlibs in R
- [dadjokeapi](#): Return a Dad Joke
- [demotivr](#): Show Demotivational Messages on Errors
- [kittyR](#): Kitty pictures and meows from R console
- [praise](#): Praise Users
- [memer](#): A tidyverse compatible package for creating memes in R using magick
- [rainbowwrite](#): Rainbow Coloured Output
- [wesanderson](#): A Wes Anderson Palette Generator

```
# Noises!
```

```
beepR::beep(2)
```

```
kittyR::meowR(4)
```

```
BRRR::skrrrahh(16)
```

```
# Messages
```

```
dadjokeapi::groan()
```

```
praise::praise()
```

```
# Images and colors
```

```
library(magrittr)
```

```
memer_get("DosEquisMan") %>%
```

```
  meme_text_top("I don't always  
useR!", size = 28) %>%
```

```
  meme_text_bottom("But when I do, I  
procrastinate\nby making memes.",  
size = 28)
```

```
joke <- dadjokeapi::groan(FALSE)$joke
```

```
rainbowwrite::lolcat(joke)
```

Statistics 1 minute demo

```
# Read data from a URL
nla_url <- "https://www.epa.gov/sites/production/files/2014-10/nla2007_chemical_conditionestimates_20091123.csv"
nla_wq <- read.csv(nla_url)

# Remove missing data
nla_wq <- nla_wq[complete.cases(nla_wq[,c("CHLA", "PTL", "NTL"))],]

# Some stats
mean(nla_wq$CHLA)
t.test(nla_wq$CHLA ~ nla_wq$LAKE_ORIGIN)
nla_lm <- lm(log10(CHLA) ~ log10(PTL) + log10(NTL),
            data = nla_wq)
summary(nla_lm)
```

Visualization 1 minute demo

```
# Use ggplot2 to create viz
library(ggplot2)

# Iteratively add parts to my plot
nla_wq %>%
  ggplot(aes(x=NTL, y=PTL)) +
  geom_point(aes(size=CHLA, color=CHLA)) +
  scale_x_log10() +
  scale_y_log10() +
  scale_color_continuous(low = "springgreen", high =
"darkgreen") +
  geom_smooth(method = "lm", color = "grey50") +
  theme_classic() +
  labs(title = "Total Nitrogen, Total Phosphorus, and
Chlorophyll Associations",
       x = "Log 10 (Total Nitrogen)",
       y = "Log 10 (Total Phosphorus)")
```

Geospatial 1 minute demo

```
# Load up several spatial packages
```

```
library(USAboundaries)
```

```
library(sf)
```

```
library(dplyr)
```

```
library(elevatr)
```

```
library(mapview)
```

```
library(raster)
```

```
# Get map of lower 48 states
```

```
usa_l48 <- us_boundaries() %>%  
  filter(state_abbr != "HI" &  
          state_abbr != "AK" &  
          state_abbr != "PR")
```

```
# Get DEM. Zoom level 4 returns DEM of ~ 3.8km resolution
```

```
usa_l48_dem <- get_elev_raster(usa_l48, src="aws", z = 4,  
                              clip = "location")
```

```
# Create quick leaflet map
```

```
mapview(usa_l48_dem) + usa_l48
```


Documents 1 minute demo

- R mixed with Markdown:
 - <https://daringfireball.net/projects/markdown>
- PDF
- Word Documents
- Presentations



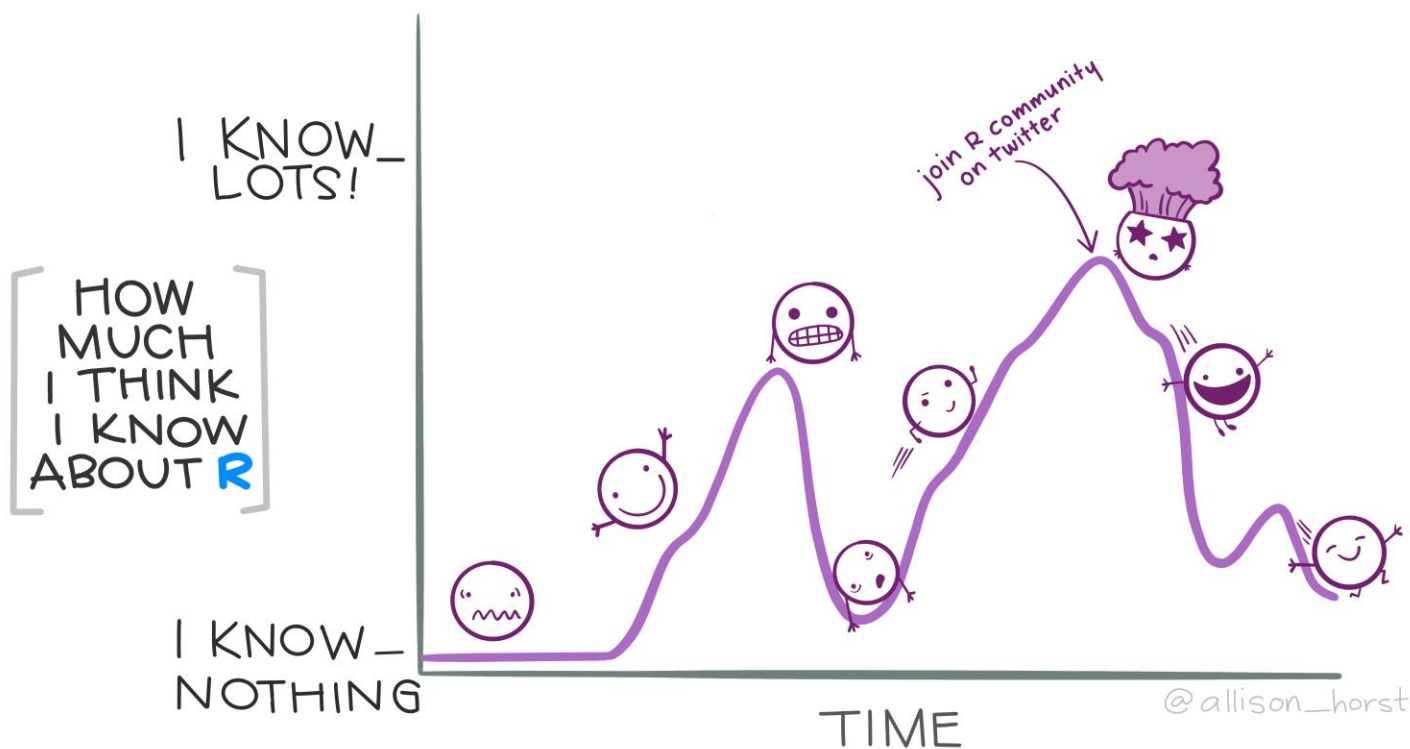
Web 1 minute demo

- Shiny
- Shiny examples:
 - <https://shiny.rstudio.com/gallery/>
- Shiny at EPA
 - <https://shiny.epa.gov/fcedts/>
 - Internal at NCC
 - Working on other options
- Web and HTML tools galore
 - Visualization
 - Web Scraping
 - APIs



Learn More

- Twitter
 - #rstats, #TidyTuesday, #rspatial
- EPA R User Group
 - Monthly meetings, face to face meetings, webinars
- learnr:
 - <https://rstudio.github.io/learnr/>
- The Carpentries:
 - <https://carpentries.org/>



Questions

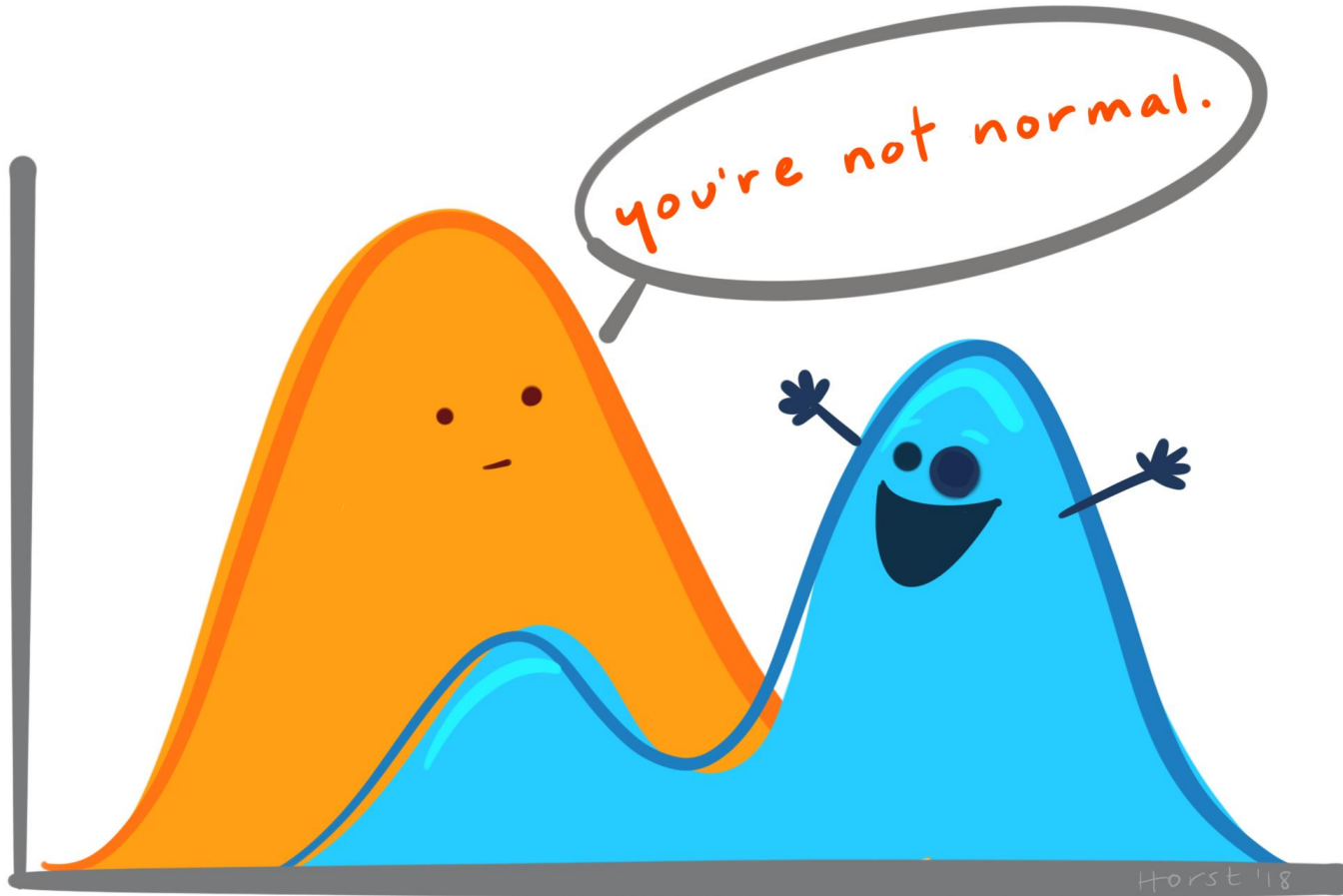


Illustration credit: Allison Horst, <https://github.com/allisonhorst/stats-illustrations>