1. Get the script in R: download.file("

   http://acis.ufl.edu/~mcollins/r_pkg_lesson.R",

   destfile="r_pkg_lesson.R")

2. Get an API token from NOAA here: http://www.ncdc.noaa.gov/cdo-web/token
3. Sign up for a plot.ly account here: http://plot.ly/
4. Install the following R packages:

   jsonlite, rnoaa, gridExtra, ridigbio, plotly,

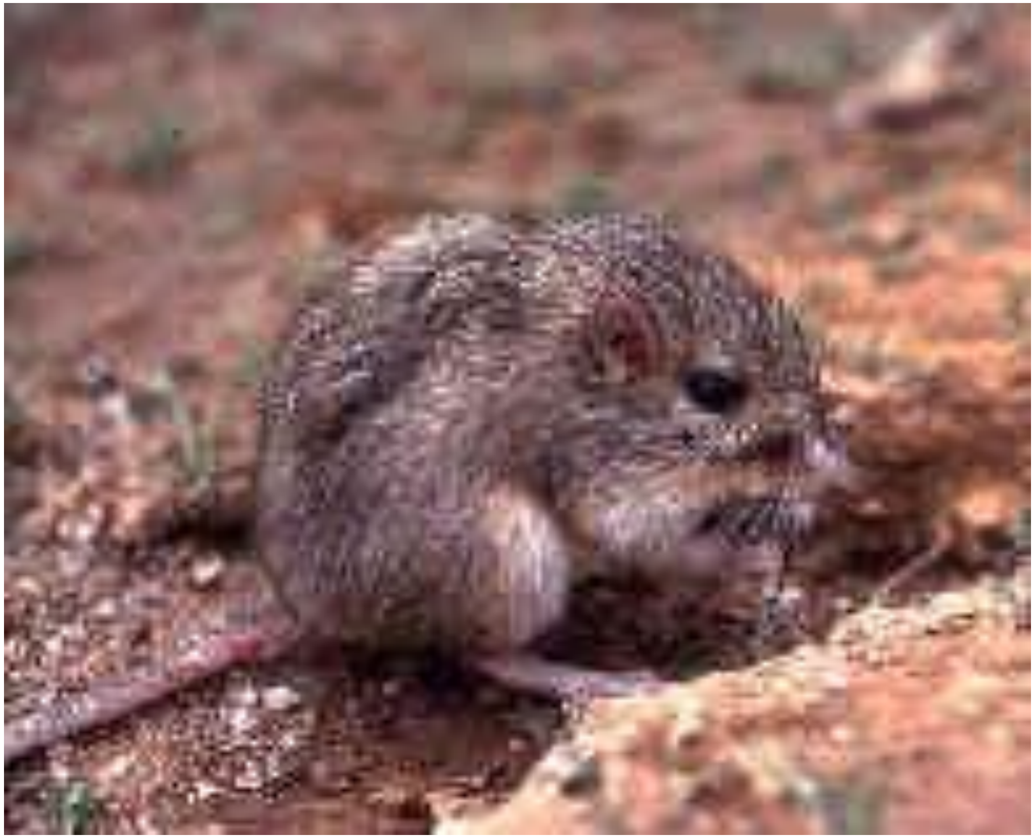
Using APIs in R

Matthew Collins

March 12, 2015

iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Desert pocket mouse (*Chaetodipus penicillatus*)



"Desert pocket mouse" by FWS - http://www.fws.gov/southwest/refuges/arizona/cabeza/index.html.
Licensed under Public Domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Desert_pocket_mouse.jpg#mediaviewer/File:Desert_pocket_mouse.jpg

Winter activity of desert pocket mice

“In winter, these pocket Mice lower their body temperature and enter a state of inactivity known as torpor. They wake occasionally to nibble on the food they have stored.”

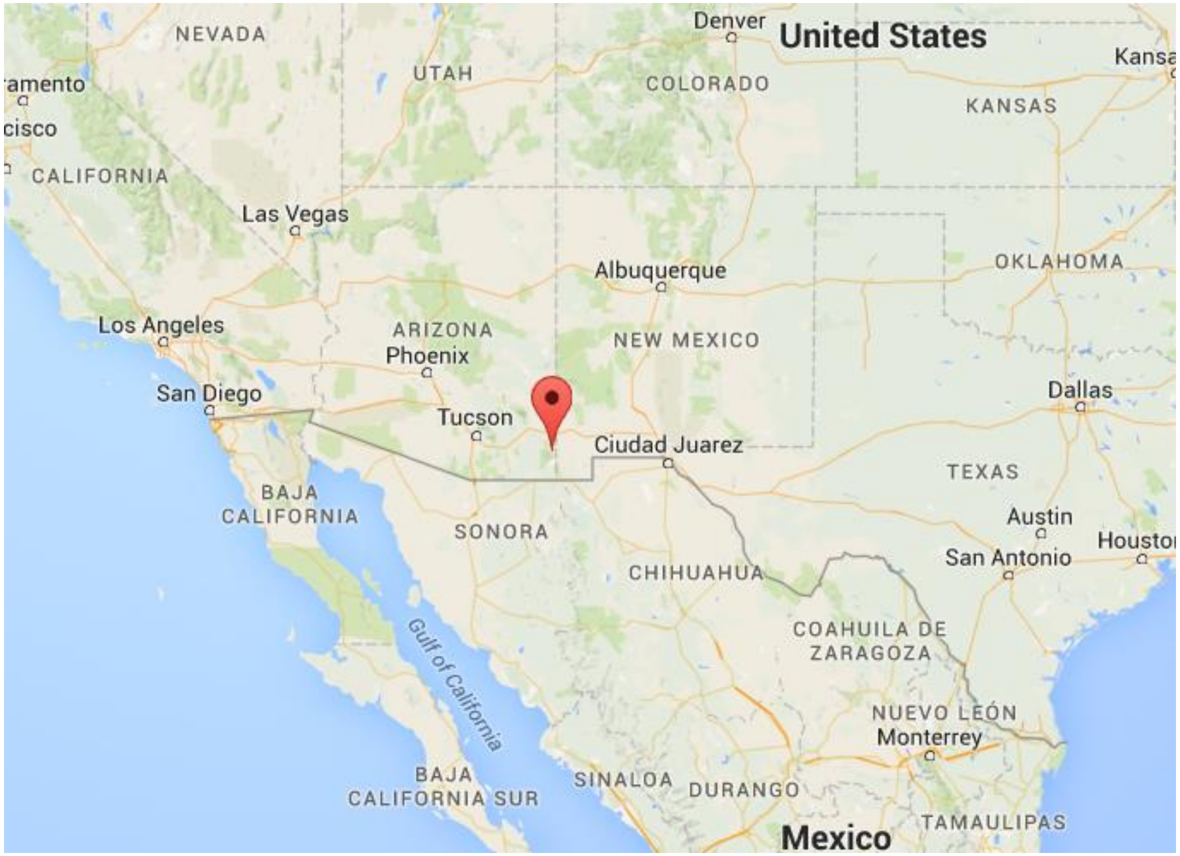
http://eol.org/pages/311995/overview

“*Chaetodipus penicillatus* may be active all year round in some areas, though it is inactive in the winter in southern Arizona.”

http://en.wikipedia.org/wiki/Desert_pocket_mouse

Can we see evidence of this in the data we have?

Where was that survey data from?

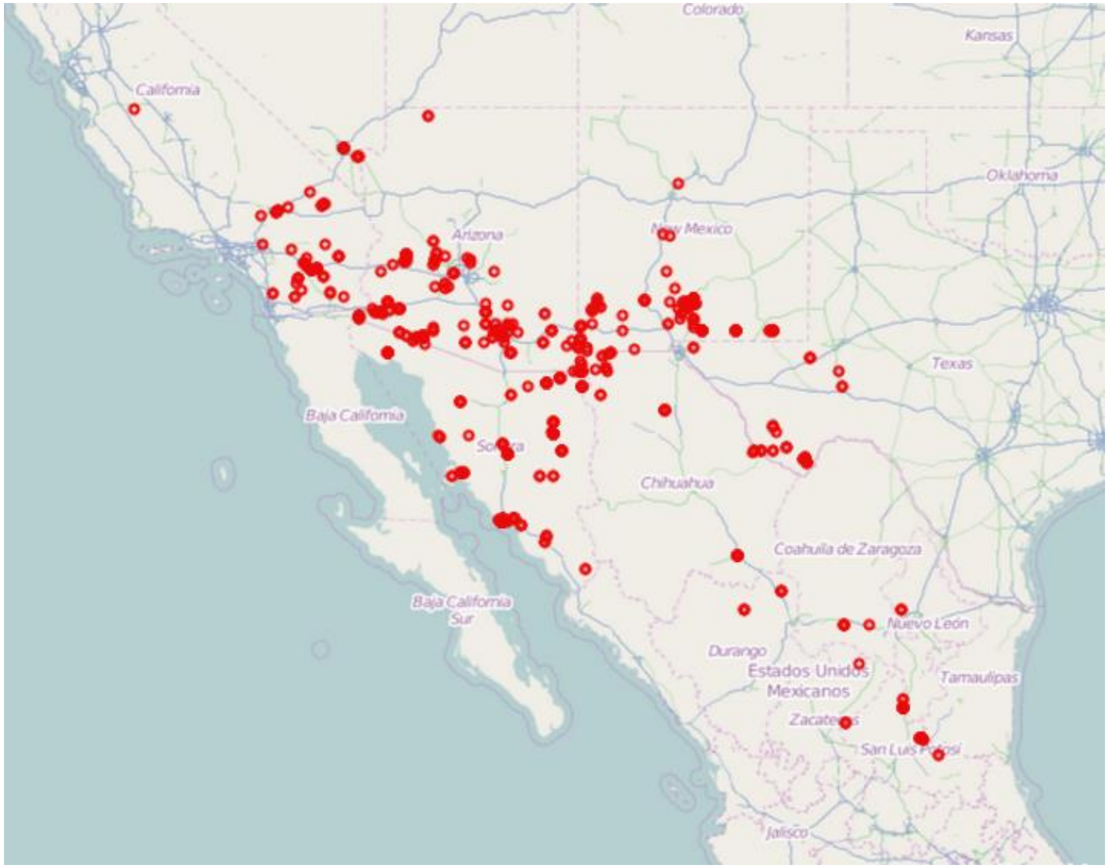


Portal in Southern Arizona

How convenient!

https://www.google.com/maps/place/Portal,+AZ+85632/@31.9137023,-109.1414495,5z/data=!4m2!3m1!1s0x86d995001c757413:0x1e02844f7993e453

Where can we get more data?



iDigBio!

(and it's not just from southern Arizona)

https://portal.idigbio.org/portal/search

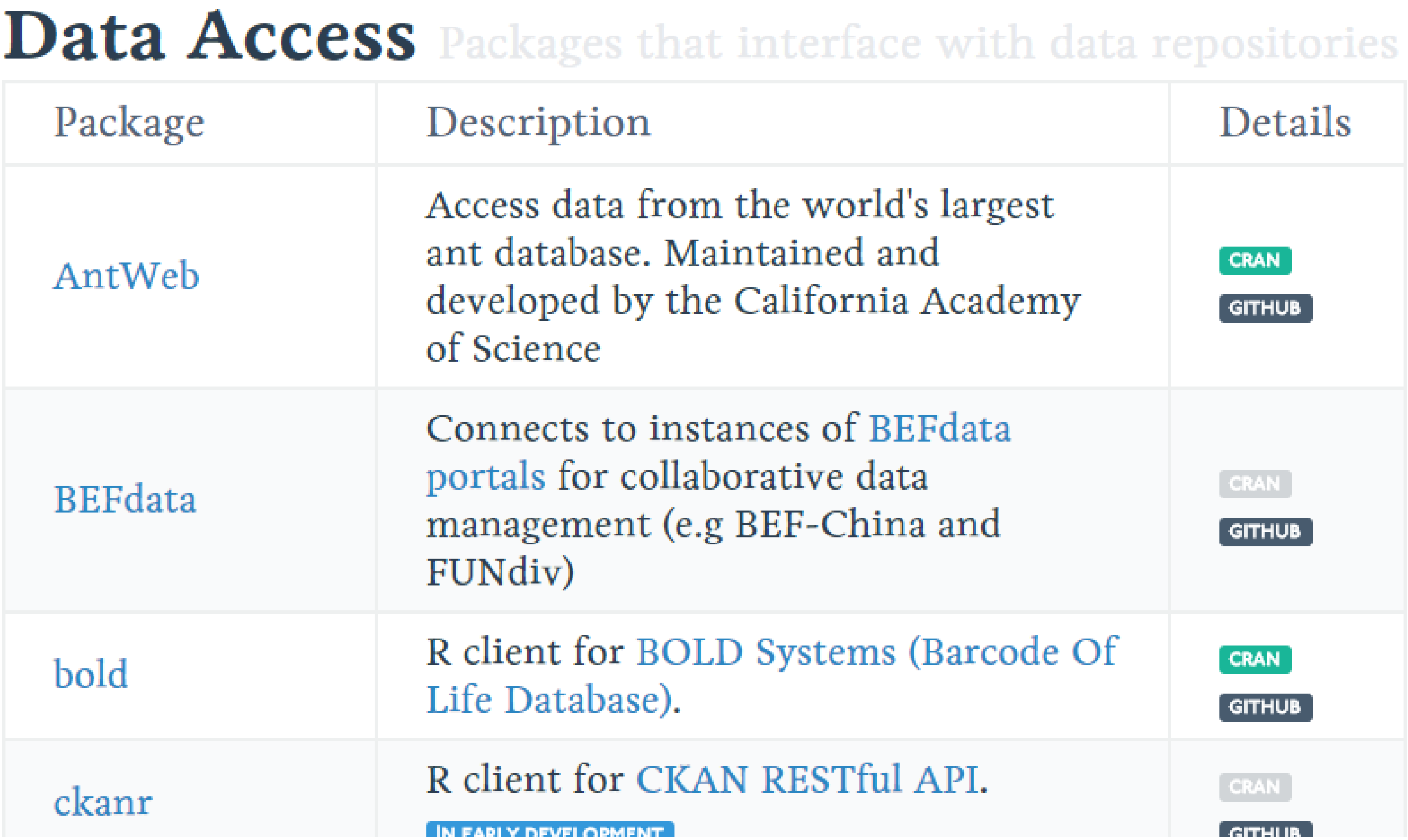
To R!

Where can we find packages?

- CRAN - The Comprehensive R Archive Network (http://cran.r-project.org/)
- rOpenSci – R Open Science (http://ropensci.org/packages/)
- GitHub – (https://github.com)
- Your colleagues!

rOpenSci is easiest

rOpenSci package listing snippet



Data Access Packages that interface with data repositories

Package	Description	Details
AntWeb	Access data from the world's largest ant database. Maintained and developed by the California Academy of Science	CRAN GITHUB
BEFdata	Connects to instances of BEFdata portals for collaborative data management (e.g BEF-China and FUNdiv)	CRAN GITHUB
bold	R client for BOLD Systems (Barcode Of Life Database).	CRAN GITHUB
ckanr	R client for CKAN RESTful API. IN EARLY DEVELOPMENT	CRAN GITHUB

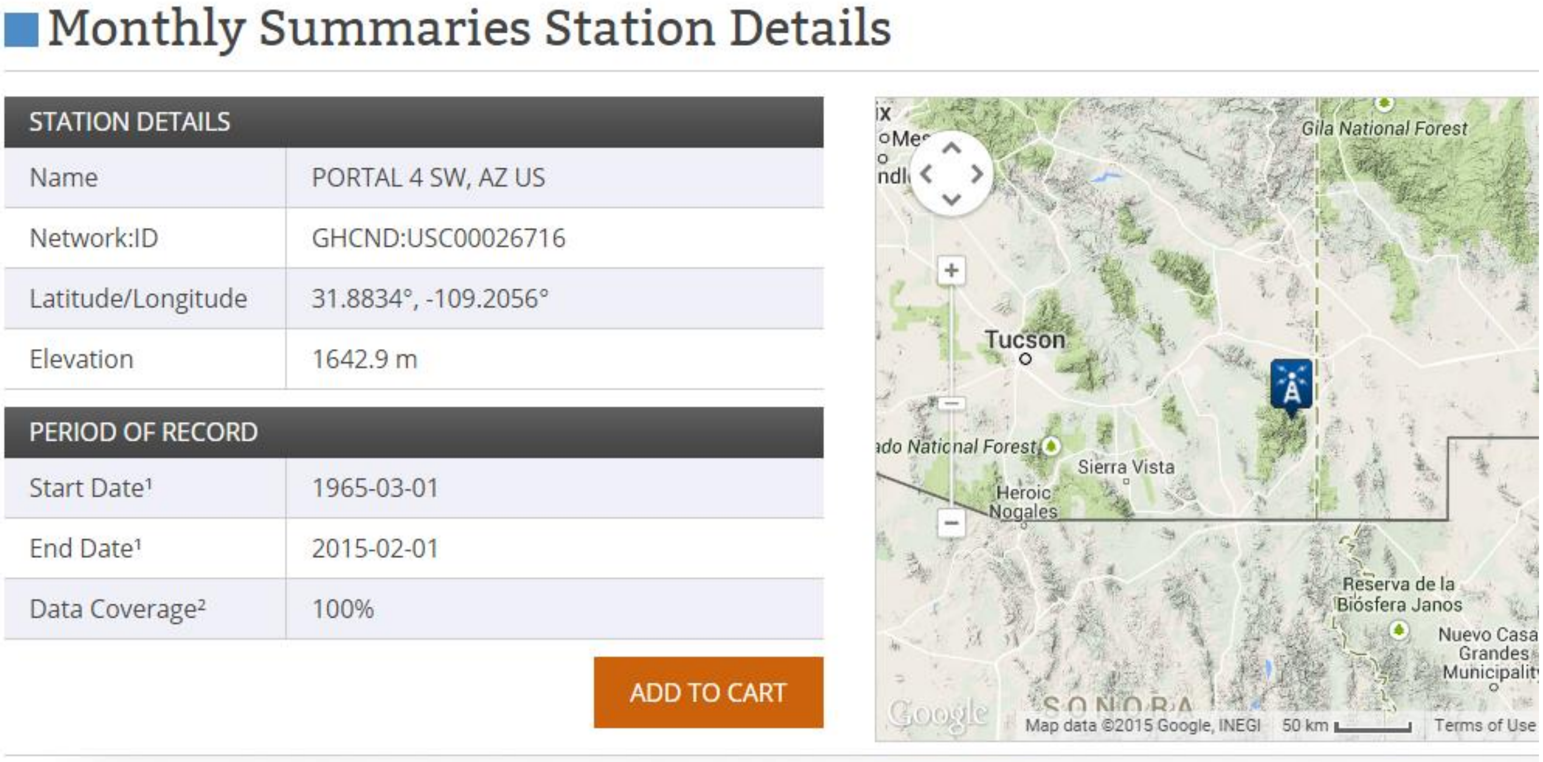
http://ropensci.org/packages/index.html

How to learn about R packages

- CRAN pages – http://cran.r-project.org/web/packages/rnoaa/index.html
- rOpenSci pages – http://ropensci.org/tutorials/rnoaa_tutorial.html
- Github - https://github.com/ropensci/rnoaa
- Vignettes – By many people
- The “?” for help in R

Getting NOAA data for Portal, AZ

http://www.ncdc.noaa.gov/cdo-web/datatools/findstation



Monthly Summaries Station Details

STATION DETAILS	
Name	PORTAL 4 SW, AZ US
Network:ID	GHCND:USC00026716
Latitude/Longitude	31.8834°, -109.2056°
Elevation	1642.9 m

PERIOD OF RECORD	
Start Date[1]	1965-03-01
End Date[1]	2015-02-01
Data Coverage[2]	100%

http://www.ncdc.noaa.gov/cdo-web/datasets/GHCNDMS/stations/GHCND:USC00026716/detail

Now that we know what station is in Portal, back to R!

tlnrvjzLToJojhXEsUYzeGNmUEUHurYk

Metadata for GHCNDMS

http://www1.ncdc.noaa.gov/pub/data/cdo/documentation/GHCNDMS_documentation.pdf

Selected quotes from metadata

Air Temperature (all units in Fahrenheit on PDF monthly form and tenths of degrees Celsius on CSV or text)

MNTM – Monthly mean temperature *

http://www1.ncdc.noaa.gov/pub/data/cdo/documentation/GHCNDMS_documentation.pdf

Back to R!

Authentication

Web APIs often need to know who you are

1. Limit rate of requests
2. Actions operate on your account
3. Collect statistics for their reporting

Types of access

- Open access like iDigBio (no user identification, no rate limits)
- Simple user identification like NOAA (rate limits, reporting)
- User account authentication like plot.ly (write and modify a user account)

Plot.ly authentication

Go to this URL:

https://plot.ly/ggplot2/getting-started/

And look for the heading Authentication

Back to R!

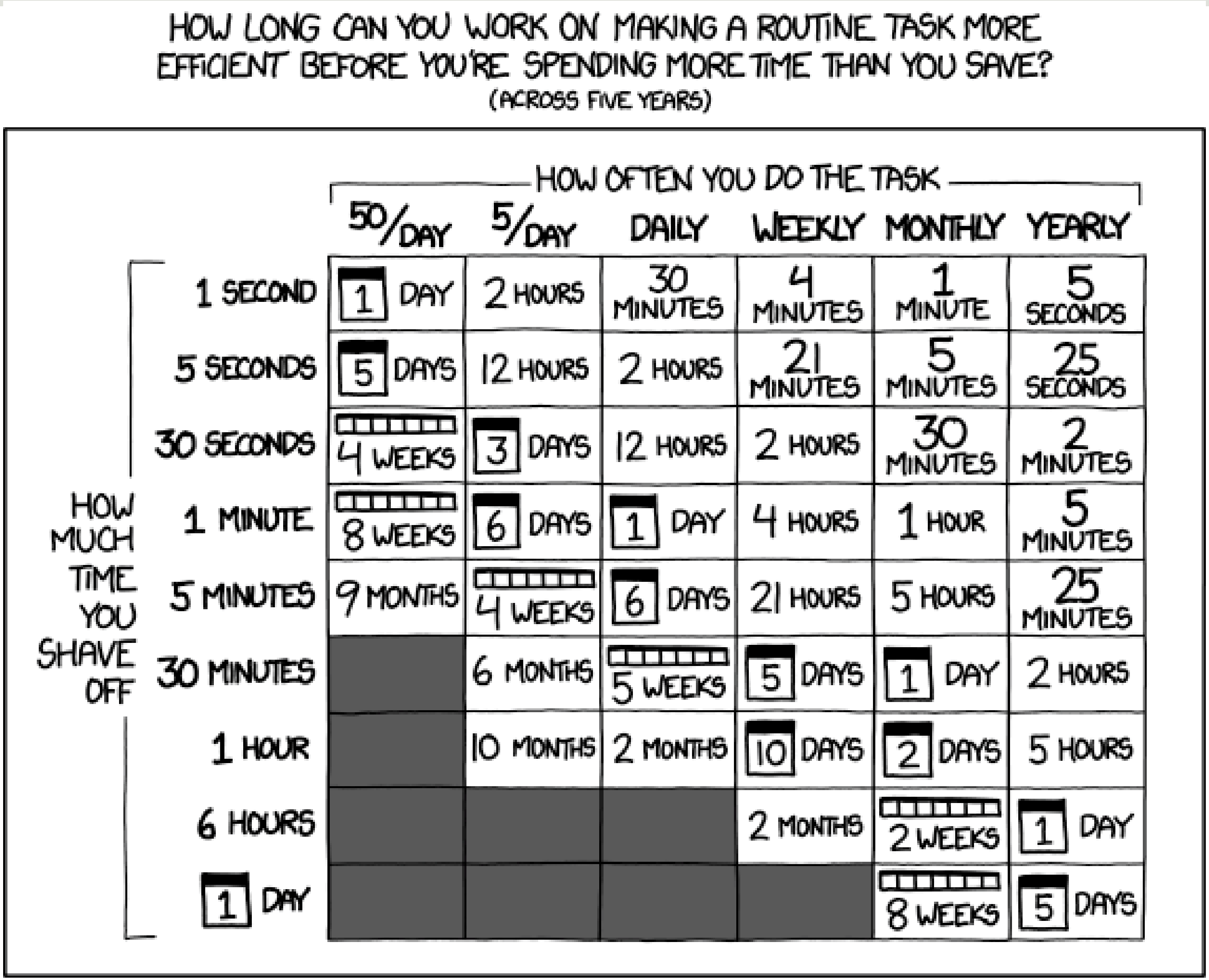
Summary

In this session we

1. Made plots from data we had as CSV files
2. Loaded data from iDigBio and plotted it
3. Loaded data from NOAA and plotted it
4. Shared our plots on plot.ly

Takeaways

1. Packages and APIs give you access to functionality and data you can’t do yourself or would take you too long to do
2. Finding good packages and APIs is hard
3. Understanding what they do is harder
4. But, re-read #1



HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?

(ACROSS FIVE YEARS)

		HOW OFTEN YOU DO THE TASK					
		50/DAY	5/DAY	DAILY	WEEKLY	MONTHLY	YEARLY
HOW MUCH TIME YOU SHAVE OFF	1 SECOND	1 DAY	2 HOURS	30 MINUTES	4 MINUTES	1 MINUTE	5 SECONDS
	5 SECONDS	5 DAYS	12 HOURS	2 HOURS	21 MINUTES	5 MINUTES	25 SECONDS
	30 SECONDS	4 WEEKS	3 DAYS	12 HOURS	2 HOURS	30 MINUTES	2 MINUTES
	1 MINUTE	8 WEEKS	6 DAYS	1 DAY	4 HOURS	1 HOUR	5 MINUTES
	5 MINUTES	9 MONTHS	4 WEEKS	6 DAYS	21 HOURS	5 HOURS	25 MINUTES
	30 MINUTES		6 MONTHS	5 WEEKS	5 DAYS	1 DAY	2 HOURS
	1 HOUR		10 MONTHS	2 MONTHS	10 DAYS	2 DAYS	5 HOURS
	6 HOURS				2 MONTHS	2 WEEKS	1 DAY
	1 DAY					8 WEEKS	5 DAYS

http://xkcd.com/1205/