Week 11: Managing resources from afar via APIs

PSY 525

Rick O. Gilmore

2020-03-31 09:42:04

Preliminaries

Check-in

Announcements

Today's topics

- Managing resources from afar via APIs

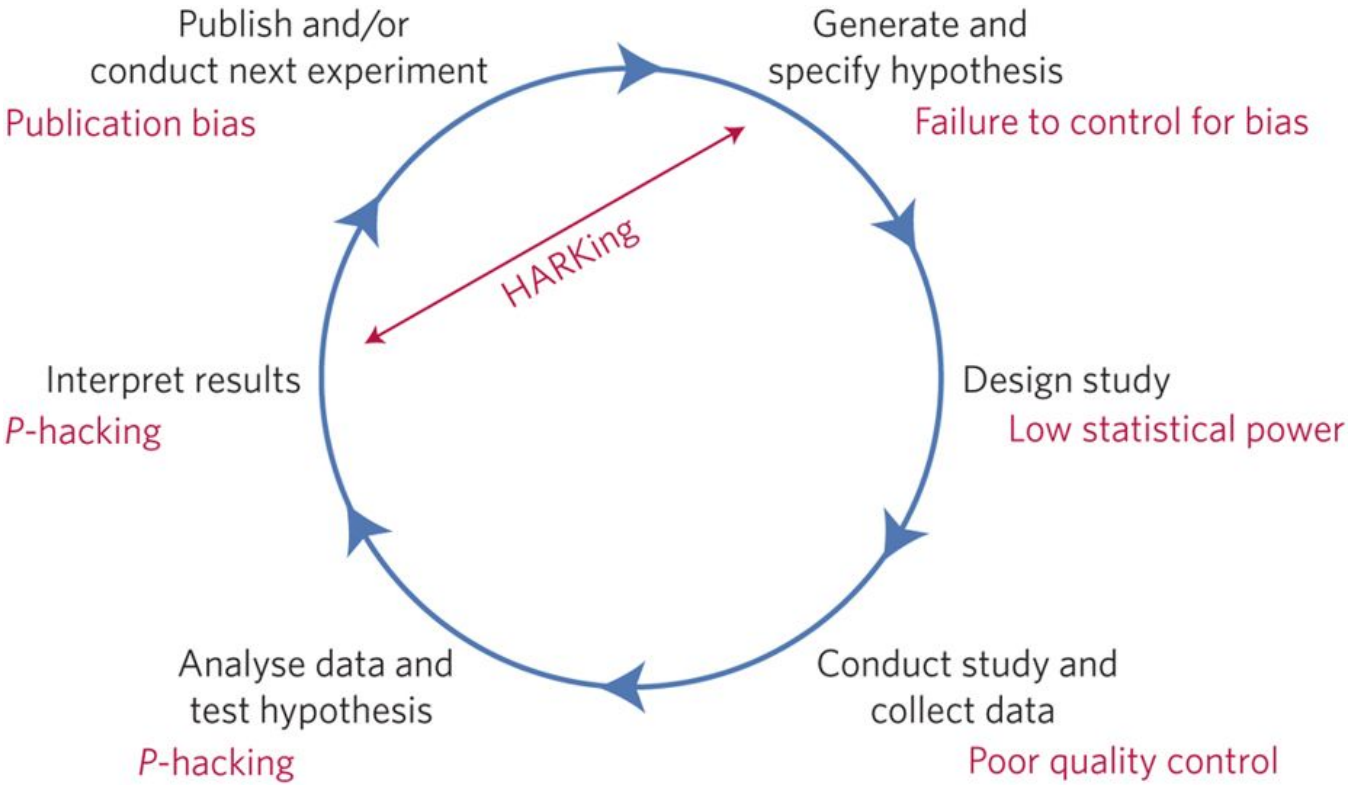
Managing resources from afar via APIs

What are APIs

- Application Program Interface (API)
- Talk to computer services

Why use APIs

- Reduce human manipulation of data files
- Reduce errors
- Reduce security/privacy breaches
- Improve reproducibility


Publish and/or
conduct next experiment
Publication bias
Generate and
specify hypothesis
Failure to control for bias
HARKing
Interpret results
P-hacking
Design study
Low statistical power
Analyse data and
test hypothesis
P-hacking
Conduct study and
collect data
Poor quality control

Munafo et al.

Some useful APIs

- Box
- Google Drive
- OSF
- Databrary

Box.com

- `boxr` R package
- `python-box` Python package

Google Drive

- `googledrive` R package
- `google-api-python-client` Python client

Open Science Framework (OSF)

- `osfr` package for R
- `osfsync` package for Python

Databrary

- `databraryapi` R package
- `databrarypi` Python package

What can you do with these APIs?

- Move data to/from cloud storage
- Leave data on cloud storage; clean, visualize locally
- Produce reproducible workflows from the get-go
- Reduce the likelihood that data can “leak”
- Access & visualize data shared by others

What's involved?

- Downloading package
- Configuring authentication
- Testing the connection
- Writing code to do what you want

Example 1: `databraryapi`

- Summarize demographics of specific study

Tamis-LeMonda, C. (2013). Language, cognitive, and socio-emotional skills from 9 months until their transition to first grade in U.S. children from African-American, Dominican, Mexican, and Chinese backgrounds. *Databrary*. Retrieved March 29, 2020 from http://doi.org/10.17910/B7CC74.

- Login

```
databraryapi::login_db()

## Please enter your Databrary user ID (email).

## Email:

## [1] FALSE
```

- Download demographic data

```
demog <- databraryapi::download_session_csv(8)
```

```
str(demog)
```

```
## 'data.frame':    1351 obs. of  36 variables:
##  $ session_id           : int  190 191 192 193 194 195 196 197 198 199 ...
##  $ session_name         : logi  NA NA NA NA NA NA ...
##  $ session_date         : chr  "2006-XX-XX" "2006-XX-XX" "2006-XX-XX" "2006-XX-XX" ...
##  $ session_release      : chr  "EXCERPTS" "PRIVATE" "EXCERPTS" "PRIVATE" ...
##  $ participant.ID       : int  3814 3811 3808 3806 3804 3803 3801 3800 3797 3796 ...
##  $ participant.birthdate: logi  NA NA NA NA NA NA ...
##  $ participant.gender   : chr  "Male" "Female" "Female" "Male" ...
##  $ participant.race     : chr  "Black or African American" "Black or African American" "Bl
##  $ participant.ethnicity: chr  "Dominican" "Dominican" "Dominican" "Unknown or not reporte
##  $ participant.language : chr  "English" "English" "English" "English" ...
##  $ group.name           : chr  "14 month" "14 month" "14 month" "14 month" ...
##  $ task1.name           : chr  "Novel toy: string with beads" "Novel toy: string with bead
##  $ task1.description    : chr  "Mother-child play with beads and string\nMother-child dyad
##  $ task2.name           : chr  "Mother-child free play" "Mother-child free play" "Mother-c
##  $ task2.description    : chr  "Child alone play with standard set of toys\nMother-child d
##  $ task3.name           : chr  "Numeracy book" "Numeracy book" "Numeracy book" "Numeracy b
##  $ task3.description    : chr  "Mother-child booksharing wordless number book\nMothers sha
##  $ task4.name           : chr  "Emotion book" "Emotion book" "Emotion book" "Emotion book"
##  $ task4.description    : chr  "Mother-child booksharing wordless book of baby faces expre
##  $ task5.name           : chr  "Familiar toy" "Familiar toy" "Familiar toy" "Familiar toy"
##  $ task5.description    : chr  "Mother-child play interaction with child favorite toy\nMot
##  $ task6.name           : chr  "" "" "" "" ...
##  $ task6.description    : chr  "" "" "" "" ...
```

- Clean and visualize

```
sex_race <- demog %>%
  dplyr::select(., sex = participant.gender,
                race = participant.race)

xtabs(formula = ~ sex + race, sex_race)
```

```
##         race
## sex          Asian Black or African American Unknown or not reported White
##              7     0                        0                       0     0
##   Female     0   112                      341                       4   209
##   Male       0   111                      410                       2   155
```

Other working examples

- Quality Assurance (QA) on PLAY
- Databrary weekly report

Example 2: NY Times data on COVID-19

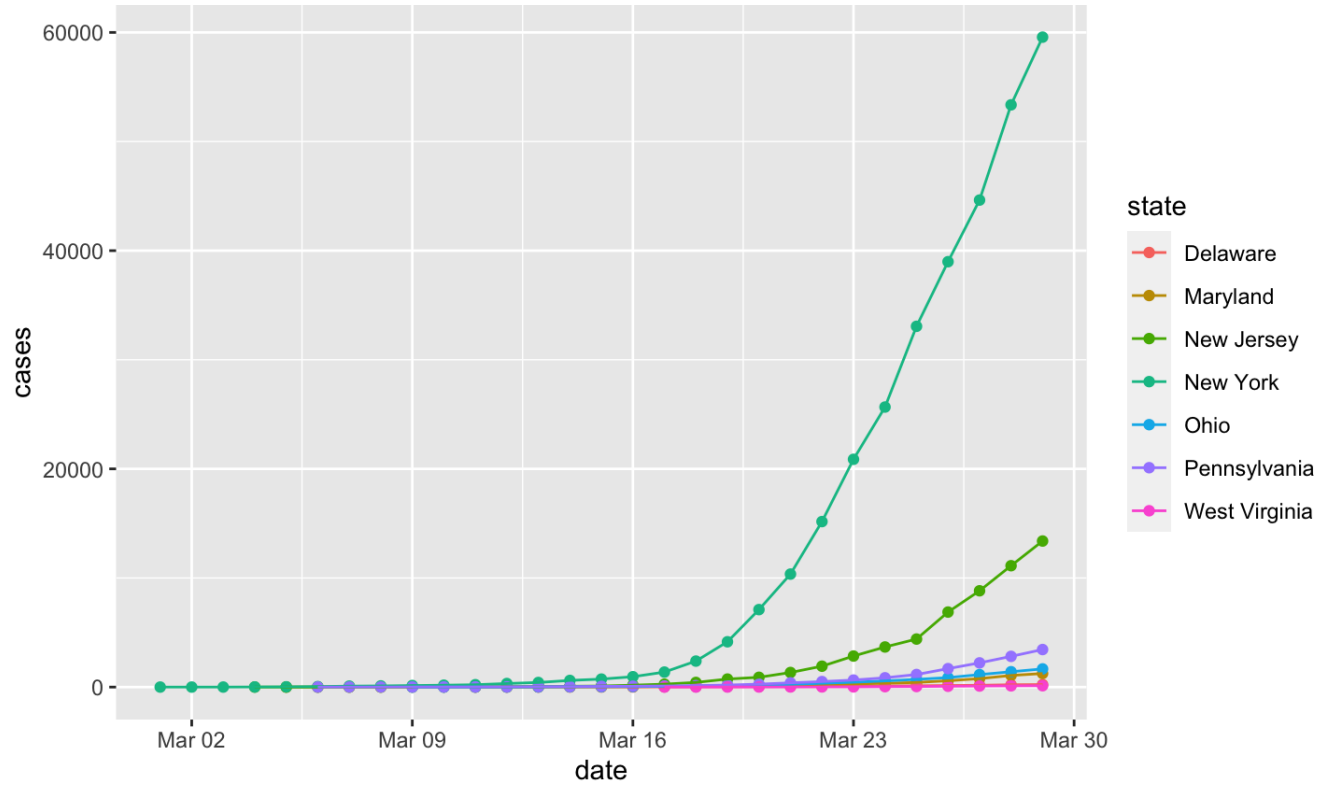
- Data: https://github.com/nytimes/covid-19-data
- NY tracking page: https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html

```
# Note the URL uses raw.githubusercontent.com
cv19 <- readr::read_csv("https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-st
```

```
## Parsed with column specification:
## cols(
##   date = col_date(format = ""),
##   state = col_character(),
##   fips = col_character(),
##   cases = col_double(),
##   deaths = col_double()
## )
```

```
str(cv19)
```

```
cv19 %>%
  dplyr::filter(., state %in% c("Pennsylvania", "New York", "New Jersey", "Maryland", "Ohio",
  ggplot(.) +
  aes(date, cases, color = state) +
  geom_point() +
  geom_line()
```



Your turn

Connect to Qualtrics

- how_to/interact-with-qualtrics

Connect to Box

- how_to/interact-with-box

Connect to Google Drive

- how_to/interact-with-google-drive

Next time…

- Where to share
- Your open science portfolio

Resources

Software

This talk was produced on 2020-03-31 in RStudio using R Markdown. The code and materials used to generate the slides may be found at https://github.com/psu-psychology/psy-525-reproducible-research-2020. Information about the R Session that produced the code is as follows:

```
## R version 3.6.2 (2019-12-12)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
```

References