

L14 Tips & Tricks

EPID 799B

Mike Dolan Fliss

Fall 2016

10.17.2016

Tips & Tricks

- Tables
- Reading multiple files
- Combining datasets
- The Hadleyverse
- With/within
- Booleans
- Regular expressions (e.g. ICDs)
- Control tables
- Stack Exchange
- Webscraping
- APIs
- Keyboard shortcuts (again!)
- Think aheading...

Tables

- RGoogleDocs package - collaboration at its finest!
- `gmodels::CrossTable()` – SAS like tables.
- `rbind` results line by line into a results table. Write to csv next to excel formatted table update table. (^ will cover later...)
- Use `read.table()` / `write.table()` and `dput()`

Reading multiple files

- Great for many datasets
 - Can read, spit out basic stats for each into a table, combine all files, etc.

```
setwd("D:/some dir/buncha files/")  
filelist = list.files(recursive = T)  
for (f in filelist){  
  ### Do something with f, like read into a data.frame....  
}
```

Combining data

- `dplyr::bind_rows()`, formerly `rbind_all()`, combines data.frames and matches column names regardless of order.
- Merge is nice in a pinch. dplyr has sql like joins.
- Merge doesn't guarantee order! (e.g. spatial data)
- In general, dplyr is worth investing in.

The Hadleyverse

... I mean tidyverse, I guess.

<https://blog.rstudio.org/2016/09/15/tidyverse-1-0-0/>

AMA:

https://www.reddit.com/r/dataisbeautiful/comments/3mp9r7/im_hadley_wickham_chief_scientist_at_rstudio_and/

No, really, subscribe to something.

Beginners

- **readr**
Reading files
- **lubridate**
Date and time handling
- **stringr**
Working with text strings
- **tidyr**
Make tidy data
- **dplyr**
Data manipulation
- **ggplot2**
The graphical grammar

Advanced

- **ggvis**
Web graphics
- **rvest**
Web scraping
- **devtools**
Package development
- **roxygen2**
Document your functions
- **testthat**
Automatic code testing

With/within

- SAS DATA step style! (But seriously...)

```
### with/within
fahrenheit_to_celcius <- function(f) (f - 32) / 1.8
airquality[c("cTemp", "logOzone", "MonthName")] <- with(airquality, list(
  fahrenheit_to_celcius(Temp),
  log(Ozone),
  month.abb[Month]
))
airquality <- within(airquality,
  {
    cTemp2 <- fahrenheit_to_celcius(Temp)
    logOzone2 <- log(Ozone)
    MonthName2 <- month.abb[Month]
  })
```

FROM: <https://www.r-bloggers.com/friday-function-triple-bill-with-vs-within-vs-transform/>

Booleans

- R is precise about missingness and Booleans, which can be an issue (but an important one!) for subsetting.

x	f	y		& AND	OR
TRUE	x	TRUE	=	TRUE	TRUE
TRUE	x	FALSE	=	FALSE	TRUE
TRUE	x	NA	=	NA	TRUE
FALSE	x	FALSE	=	FALSE	FALSE
FALSE	x	NA	=	FALSE	NA
NA	x	NA	=	NA	NA

Regular Expressions

- Worth learning! A little goes a long way.

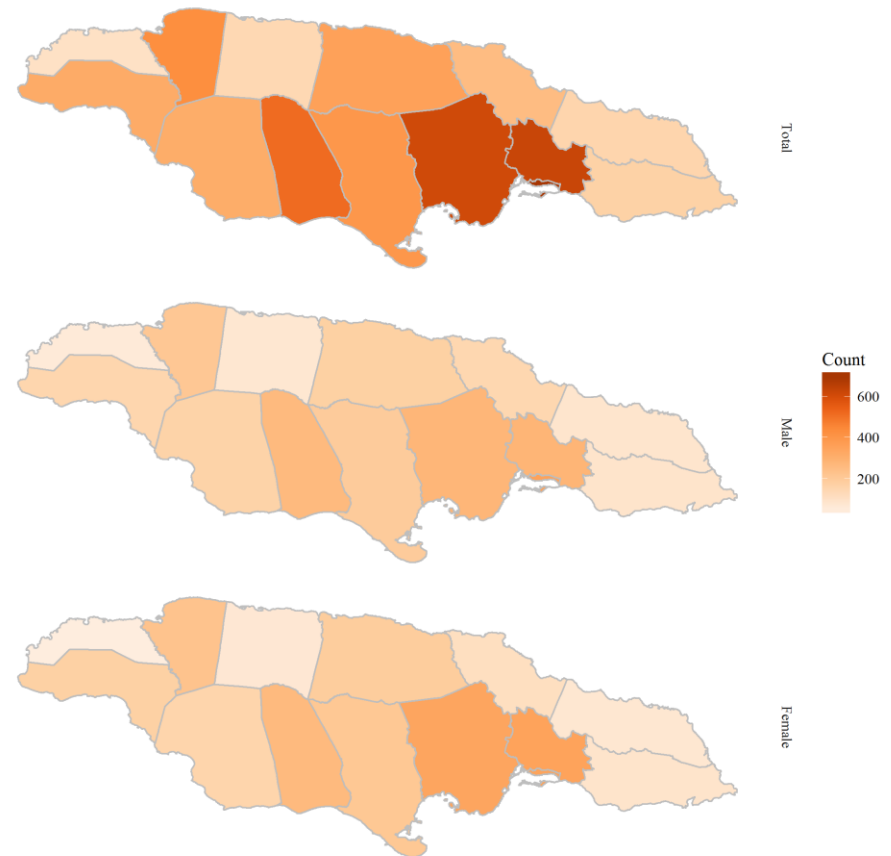
```
grepl("38[1-9]", diagnosis.df[1,])

hascode = function(df, pattern, useperl=F){
  as.integer(apply(df, 1, function (x) any(grepl(pattern, x, fixed=!useperl, perl=useperl))))
}

# Specified
d.df$F995.53 = as.integer(hascode(diagnosis.df, "995.53")) #child abuse sexual
d.df$FE960.1 = as.integer(hascode(injury.df, "E960.1")) #injury by rape
# Suggestive
d.df$F054.1 = as.integer(hascode(diagnosis.df, "54.1")) #genital herpes
d.df$F098 = as.integer(hascode(diagnosis.df, "098")) #gonococcal infection
#....
d.df$F863.1_2_3 = as.integer(hascode(diagnosis.df, "863\\. [123]", T))
```

Count of cardiovascular disease deaths by parish and gender in 2011
 n = 5,099 cardiovascular disease deaths with age and gender data present. Rate /10,000 deaths.

Control Tables



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Normal

Page Break Preview

Page Layout

Custom Views

Stack Exchange

- How to ask for help:
<http://stackoverflow.com/help/how-to-ask>
- How to create a MCV (minimal, complete, verifiable) example*:
<http://stackoverflow.com/help/mcve>
- Semi-OK Example:
<http://gis.stackexchange.com/questions/148398/how-does-spatial-polygon-over-polygon-work-to-when-aggregating-values-in-r>

Web scraping

- Sometimes really helpful



The screenshot shows the CEPH (Council on Education for Public Health) website. The header includes the CEPH logo and name, a search bar, and an 'About' button. The main heading is 'Accredited Schools & Programs'. Below this, there are links to 'Schools of Public Health' and 'Accredited Schools & Programs'. The 'Schools of Public Health' section lists several institutions with their contact information and accreditation dates. The 'Accredited Schools & Programs' section lists various programs and their accreditation dates. The right sidebar contains a 'Newsletter' section with links to '2016 revised criteria', 'Accreditation Criteria and Procedures', 'Accredited Schools & Programs', 'Applicants', 'Search for a Degree Program', 'Application Templates', 'CEPH Contact: Brittney Lilly', 'CEPH Criteria Rollout Meeting: Tentative Agenda', 'Criteria Rollout Meeting: Oct. 30, 2016', 'Data Templates', 'Dates to Remember', 'FAQs and Technical Assistance', 'Frequently Asked Questions', 'Featured Institutions', 'Questions about the Rollout Meeting', 'Alasha O'Connell, MPH', 'Yessenia Castillo, MPH', 'Benefits of Accreditation', 'Resources for Students', 'Students' Frequently Asked Questions', 'Considering Accreditation', 'Preparing for an Accreditation Review', 'Reports to CEPH', 'Technical Assistance and Resources', 'For Students', 'Programs – Baccalaureate Level', 'Related Links', 'Schools and Programs', 'History and Overview of Baccalaureate Accreditation', 'SBP Data Templates', 'Standalone Baccalaureate Programs – Frequently Asked Questions', 'Nominate a Site Visitor', 'Resources for Site Visitors', 'Opportunity to comment on CEPH's proposed criteria revisions: Round 3', and 'Presentations and Webinars'.

Boston University
School of Public Health
715 Albany St.
Boston, MA 02118
(617) 638-4640
Dean: Sandro Galea, MD, MPH, DrPH
(1/10/1981) 12/31/2018

Brown University
MPH Program
Public Health Program
Box G-S121
Providence, RI 02912
(401) 863-2059
Dean: Fox Wetle, MS, PhD
(6/1/2002) 7/1/2021

Colorado School of Public Health
University of Colorado Denver
University of Northern Colorado
Colorado State University
13001 East 17th Avenue
Campus Box 6-119
Aurora, CO 80045
(303) 724-4448
Dean: David C. Goff, MD, PhD
(10/16/2010) 7/1/23

Ohio State University
College of Public Health
250 Cunz Hall
1841 Neil Avenue
Columbus, OH 43210-3150
(614) 247-8196
Dean: William J. Martin II, MD
(9/19/1985) 12/31/2016

Oregon State University
College of Public Health and Human Sciences
123 Women's Building
Corvallis, OR 97331
(541) 737-3220
Dean: Tammy Bray, PhD
(10/22/1996) 7/1/2019

Saint Louis University
College for Public Health and Social Justice
3545 Lafayette Ave., Suite 300
St. Louis, MO 63104-1314
(314) 977-6100
Dean: Collins O. Althibebawa, PhD, MPH
(9/23/1983) 7/1/23

San Diego State University
Graduate School of Public Health

University of Iowa
College of Public Health
105 River Street, S160 CPHB
Iowa City, IA 52242
(319) 384-5452
Dean: Susan J. Curry, PhD, MA
(10/7/2009) 7/1/2018

University of Kentucky
College of Public Health
111 Washington Avenue, Suite 112
Lexington, KY 40536-0003
(859) 218-2247
Dean: Donna K. Arnett, PhD, MSPH
(6/11/2005) 12/31/2017

University of Louisville
School of Public Health and Information Sciences
485 East Gray Street
Louisville, KY 40202
(502) 852-3299
Dean: Craig H. Blakey, PhD, MPH
(10/6/2007) 7/1/2020

University of Maryland
School of Public Health
2242 SPH Building

```
site="http://ceph.org/accredited/" #programs
install.packages("rvest")
require(rvest)
schoolpage = read_html(site)
schoolnames = schoolpage %>% html_node("#content") %>% html_nodes("h4") %>%
html_text()
schools = data.frame(schoolnames, stringsAsFactors = F)
str(schools)
write.csv(file="PHschools.csv", schoolnames, row.names = F)
shell.exec("PHschools.csv")
```

APIs

The Google Maps web services are a collection of HTTP interfaces to Google services providing geographic data for your maps applications. This guide serves only to introduce the web services and host information common to all of the different services. Individual documentation for each service is located below:

- [Google Maps Directions API](#)
- [Google Maps Distance Matrix API](#)
- [Google Maps Elevation API](#)
- [Google Maps Geocoding API](#)
- [Google Maps Geolocation API](#)
- [Google Maps Roads API](#)
- [Google Maps Time Zone API](#)
- [Google Places API Web Service](#)

Contents

- Find out which API you need
- More about the web services
- What is a web service?
- SSL Access
- Building a Valid URL
- Polite Use of Google APIs
 - Exponential Backoff
 - Synchronized Requests
- Processing Responses
 - Processing XML with XPath
 - Processing JSON with Javascript
- The sensor Parameter

- Fancy! Useful!
- Census geocoding API:
https://geocoding.geo.census.gov/geocoder/Geocoding_Services_API.pdf
- In some cases there are already wrap-around packages (e.g. tigris)
- Use cases: tobacco retailers, geocoding

Keyboard shortcuts (again!)

- Control
- Control+(Shift+)+1,2,3,4
- Control+Shift+F10 to restart



Thinking ahead...

- Google...and SAS (!) use R.
- Health depts / services transitioning to R.
- But: punch-cards is to SAS as R is to...?
- Opensource/etc. is fundamentally a different framework... but matches PhD life.