

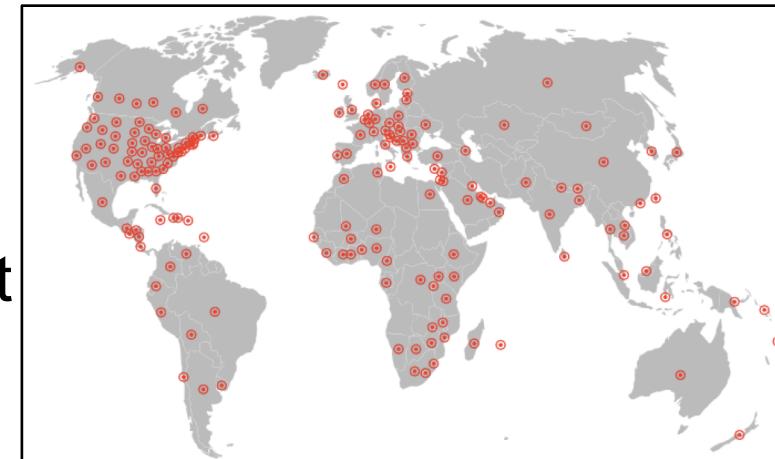
From REDCap Data to an NIH Enrollment Report

Peter D.R. Higgins (and Will Beasley)



What is REDCap?

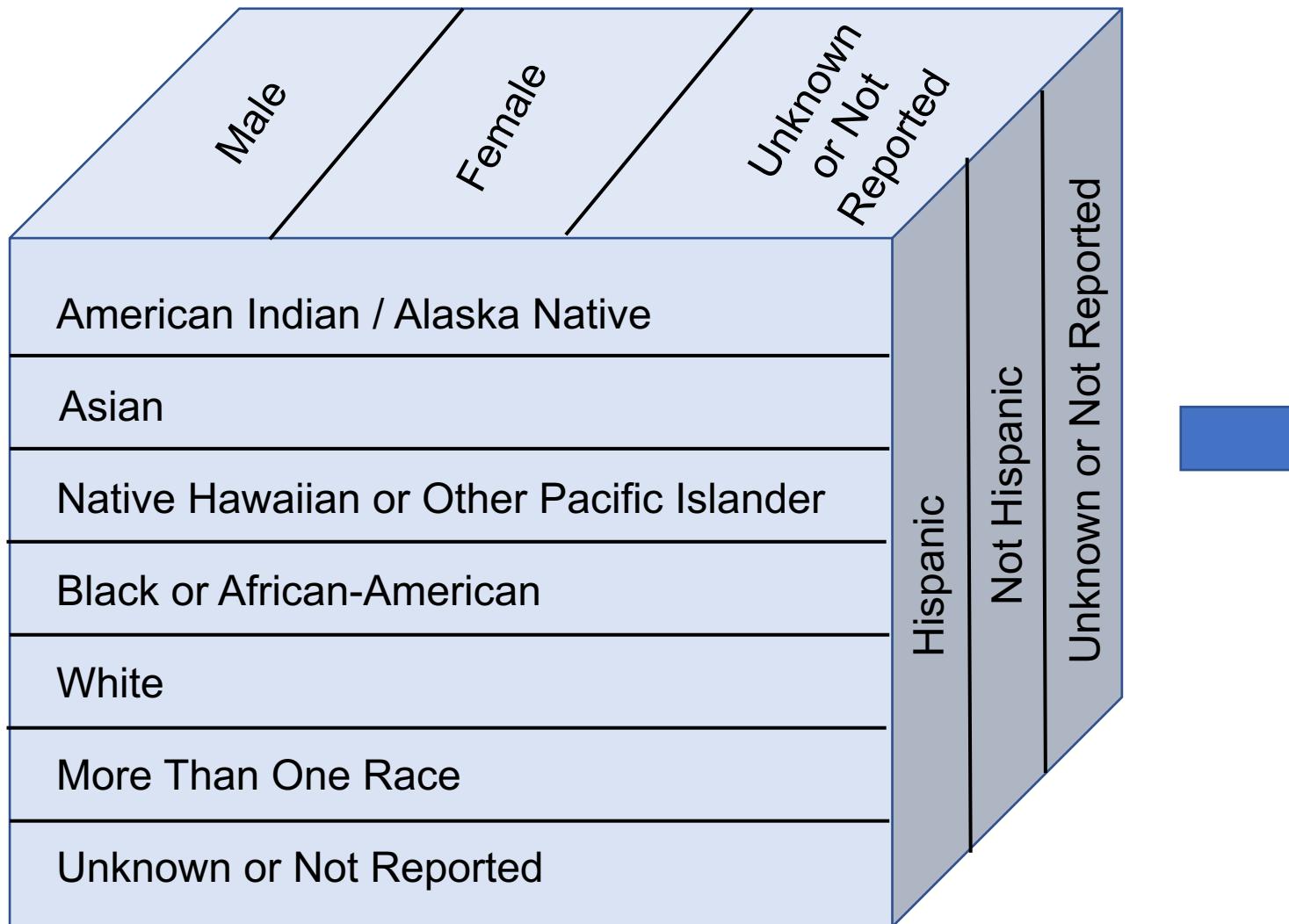
- Research Electronic Data Capture
- HIPAA compliant web database
 - Health Insurance Portability and Accountability Act
 - PHI (Personal Health Information) is protected
- Enables secure data entry from multiple sites
 - Survey data from participants
 - Data validation by field (i.e. phone #s)
- Allows real-time tracking of enrollment
- Allows real-time tracking of data collection



Used in 3,026 institutions in
126 countries worldwide
577,000 projects
777,000 users
5,697 publications
Exports data to R
Data access via API

What is an NIH Enrollment Report?

- When you have an NIH grant that includes enrolling patients
- Every year you have to provide a standard enrollment report in your annual research productivity progress report (RPPR)
- This requires you to count enrolled subjects and divides them on three dimensions
 - Race: 7 categories
 - Ethnicity: 3 categories
 - Sex: 3 categories
- To produce a 3 dimensional matrix of 63 cells of counts
- Then submit this as a 2 dimensional (very untidy) table
- With totals for both rows and columns



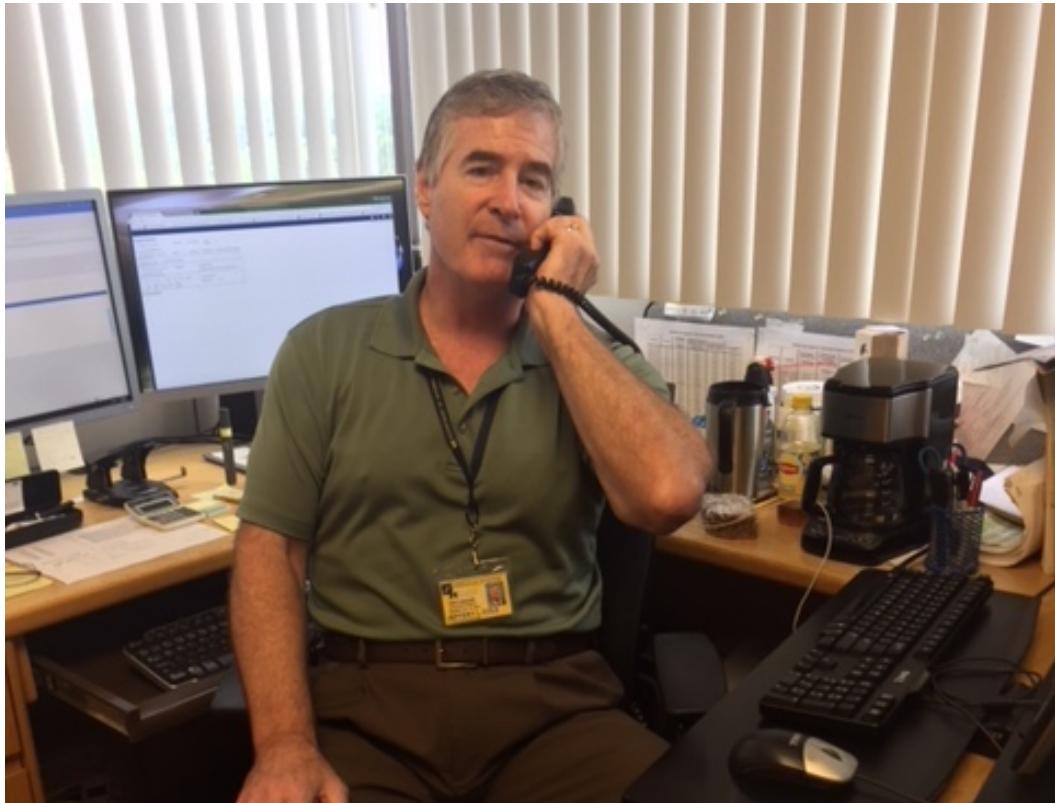
63 cell table

	Ethnic Categories								Total
	Not Hispanic or Latino			Hispanic or Latino			Unknown/Not Reported Ethnicity		
Racial Categories	Female	Male	Unknown/ Not Reported	Female	Male	Unknown/ Not Reported	Female	Male	Unknown/ Not Reported
American Indian/Alaska Native	1	0	1	0	0	0	0	0	2
Asian	1	0	0	0	0	0	0	0	1
Native Hawaiian or Other Pacific Islander	0	0	0	0	0	0	0	0	0
Black or African-American	1	1	0	0	1	0	0	0	2
White	10	11	0	2	1	0	1	2	27
More than One Race	2	3	0	1	0	0	1	1	8
Unknown or Not Reported	2	1	0	0	0	0	0	0	3
Total	17	16	1	3	2	0	1	3	0

Format of an NIH Enrollment Report

	Ethnic Categories										
	Not Hispanic or Latino			Hispanic or Latino			Unknown/Not Reported Ethnicity			Total	
Racial Categories	Female	Male	Unknown/ Not Reported	Female	Male	Unknown/ Not Reported	Female	Male	Unknown/ Not Reported		
American Indian/Alaska Native	1	0	1	0	0	0	0	0	0	2	
Asian	1	0	0	0	0	0	0	0	0	1	
Native Hawaiian or Other Pacific Islander	0	0	0	0	0	0	0	0	0	0	
Black or African American	1	1	0	0	1	0	0	0	0	2	
White	10	11	0	2	1	0	1	2	0	27	
More than One Race	2	3	0	1	0	0	1	1	0	8	
Unknown or Not Reported	2	1	0	0	0	0	0	0	0	3	
Total	17	16	1	3	2	0	1	3	0	43	

Actual Scenario



- Hi Peter.
- Do you remember when you got some NIH funds from the Peptide Center grant for the IBD databank?
- The annual progress report is due tomorrow, and it turns out that we need an NIH Enrollment Report for this. Could you get this filled out by 5 PM today?

What Normally Happens Next



REDCap Data Access

- Fake clinical trial demographics dataset hosted at OUHSC Biomedical and Behavioral Methodology Core
- Can directly access from R:



William Beasley

```
fake_redcap_demodata <- REDCapR::redcap_read_oneshot(  
  redcap_uri = "https://bbmc.ouhsc.edu/redcap/api/",  
  token      = "F304DEC3793FECC3B6DEEFF66302CAD3"  
)$data
```

- For real trial data, I would store/retrieve the token with ***keyring***

Demographic Data from REDCap

	record_id	name_last	name_first	address	phone	dob	ethnicity	race	gender	height	weight
1	1	Cornel	Alice	88 Dawnview Way	3364812635	1991-05-13	1	4	0	176.1	105
2	2	Wojdyla	Jaida	69 Jessie West St	5233183616	1931-10-14	1	4	0	173.7	127
3	3	Stunkard	Allen	49 Sibley Rd	8548788327	1964-09-15	1	3	1	170.0	166
4	4	Cimochowski	Isabell	18 Warner Pl	3986375914	1969-05-03	1	4	1	160.5	98
5	5	Woodland	Amber	43 Ardath Ct	4487462273	1969-11-22	1	4	0	183.2	112
6	6	Pistoia	Kayden	37 Cole St	4727418365	1943-08-26	1	4	1	176.8	125
7	7	Ahlquist	Gianna	98 Treat Ave	4398321526	1944-01-22	1	4	0	177.0	133
8	8	Ashlin	Chad	56 Ship St	2856641963	1931-04-14	1	4	1	182.4	144
9	9	Djokovic	Simon	44 Ironwood Way	4632973167	1962-07-07	1	1	1	160.5	100
10	10	Mosses	Zack	18 Tara St	5397286264	1978-06-20	1	3	0	147.9	118
11	11	Lucini	Magdalena	49 Conrad St	9933768739	1932-05-05	1	4	0	170.2	94

Problems : Unnecessary variables for NIH table
 Need to decode ethnicity, race, gender categories

Wrangling with dplyr & magrittr

```
fake_redcap_demodata %>%  
  filter(redcap_event_name == "baseline_visit_arm_1") %>%  
  select(sex, race:ethnicity) %>%  
  mutate(race = case_when(  
    .$race == 1 ~ "White",  
    .$race == 2 ~ "Black or African-American",  
    .$race == 3 ~ "Asian",  
    .$race == 4 ~ "Native Hawaiian or other Pacific Islander",  
    .$race == 5 ~ "American Indian or Alaska Native",  
    .$race == 6 ~ "More Than One Race",  
    .$race == 7 ~ "Unknown or Not Reported",  
    TRUE ~ "Unknown or Not Reported")) %>%
```

Wrangling with dplyr & magrittr

```
mutate(ethnic_cat = case_when(  
  .\$ethnicity == 1 ~ "Hispanic or Latino",  
  .\$ethnicity == 0 ~ "Not Hispanic or Latino",  
  TRUE ~ "Unknown or Not Reported Ethnicity")) %>%  
mutate(sex2 = case_when(  
  .\$sex == 1 ~ "Male",  
  .\$sex == 0 ~ "Female",  
  TRUE ~ "Female")) %>%  
select(sex2, race, ethnic_cat) ->  
data
```



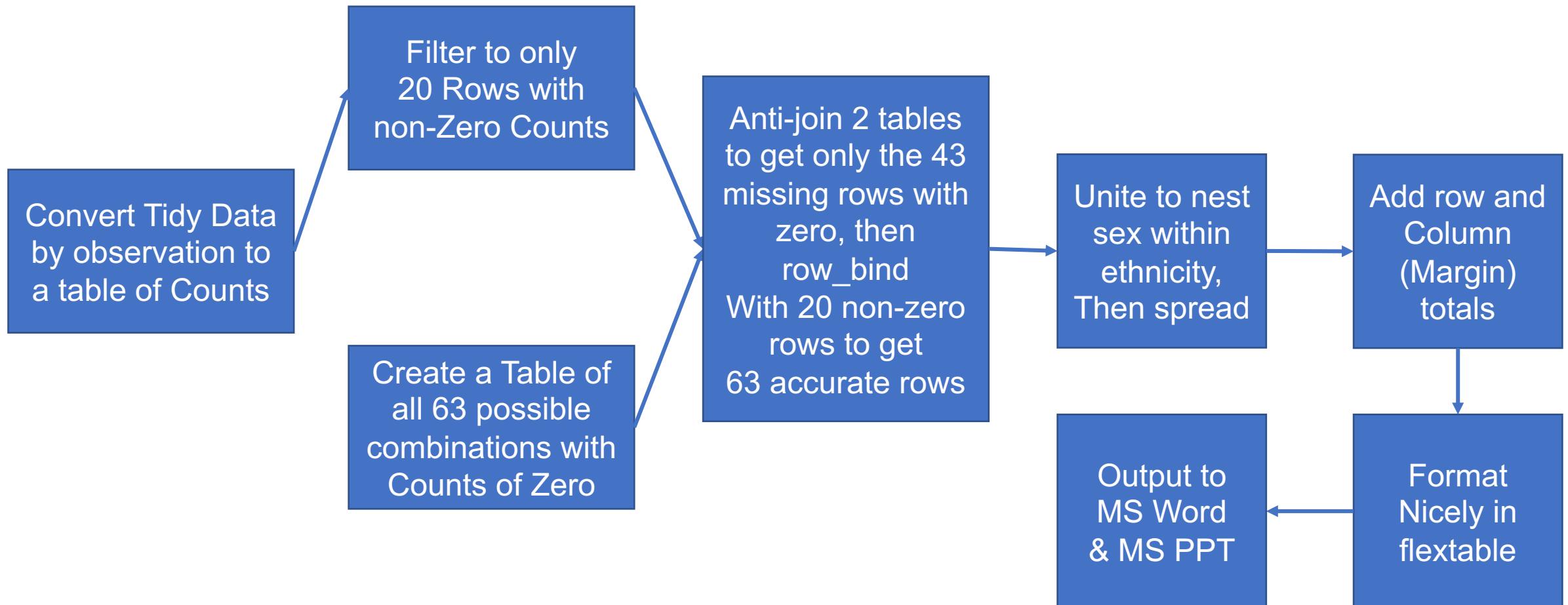
**Rightward
Assignment
Arrows
Give me
Heartburn**

Now you have tidy data

	sex2	race	ethnic_cat
1	Female	White	Hispanic or Latino
2	Female	White	Not Hispanic or Latino
3	Male	White	Not Hispanic or Latino
4	Female	White	Not Hispanic or Latino
5	Female	White	Not Hispanic or Latino
6	Male	More Than One Race	Not Hispanic or Latino
7	Male	White	Not Hispanic or Latino
8	Female	White	Not Hispanic or Latino
9	Male	White	Not Hispanic or Latino
10	Female	White	Not Hispanic or Latino
11	Male	White	Unknown or Not Reported Ethnicity

Problems : Need to convert from single observation per row to counts of all combinations.
Lots of missing combinations, i.e. Asian/Hispanic/Female – need 63 cells.
These need to be present, and need to be assigned a count of zero.

Data Wrangling Plan



Details of code slides here

- Code and live demo with fake data were previously here
 - Moved to backup slides
 - Plans changed
 - Output to MS Word 

Racial Categories	Ethnic Categories Divided by Sex										Total	
	Hispanic or Latino			Not Hispanic or Latino			Unknown/Not Reported					
	Female	Male	Unknown	Female	Male	Unknown	Female	Male	Unknown			
American Indian or Alaska Native	0	0	0	0	0	0	0	0	0	0	0	
Asian	22	33	0	2	1	0	1	0	0	59		
Black or African-American	0	0	0	0	0	0	0	0	0	0		
More Than One Race	8	10	0	0	0	0	1	0	0	19		
Native Hawaiian or Other Pacific Islander	0	0	0	0	0	0	0	0	0	0		
Unknown or Not Reported	35	29	0	3	1	0	1	1	0	70		
White	172	154	0	14	6	0	1	5	0	352		
Total	237	226	0	19	8	0	4	6	0	500		

Wrap Into A Function

- `make_nih_table(redcap_uri, api_token)`
- This works well **IF** you use the ***default*** REDCap demographics table
 - Laziness is sometimes rewarded.
 - Creativity (in this case) is punished.
- Code at https://github.com/higgi13425/nih_enrollment_table
- Feel free to fork, improve, package this code.

ORIGINAL ENDING
OF TALK

Then #rstats Collaboration Happened

- Will asked if he could build this into a proper, generalizable package
 - Input from REDCap, Forte, Medidata, others
 - Output to NIH, other standard reporting tables
- Will is the developer/maintainer of ***REDCapR***
- ***codified*** package will be on CRAN next week



William Beasley



CRAN vacation from Sep 1 to Sep 9, 2018.
During this time, the submission of packages is not possible.
Sorry for any inconvenience



Then #rstats Collaboration Happened

- Will asked if he could build this into a proper, generalizable package
 - Input from REDCap, Forte, Medidata, others
 - Output to NIH, other standard reporting tables



William Beasley

- **codified** package will be on CRAN next week
- Dev version at `devtools::install_github('OuhscBbmc/codified')`
- Main function:
`table_nih_enrollment(df)`
- Adds pretty HTML, LaTEX output with `kableExtra`

codified

Produce standard/formalized demographics tables

Clinical researchers often need to document and report the demographics of all consented subjects. A common use case is for reporting to funding agencies, including the National Institutes of Health (NIH). The NIH requires a particular format for the [PHS Inclusion Enrollment Report](#) in each annual Research Performance Progress Report ([RPPR](#)).

The [codified](#) package, in combination with the [REDCapR](#) package, provides a pipeline to directly extract the demographics of consented subjects from a REDCap database, and to rapidly and reproducibly produce standard demographics tables.

Installation and Documentation

In the future, the *release* version can be installed from [CRAN](#).

```
# install.packages("codified")
```

The *development* version can be installed from [GitHub](#) after installing the `devtools` package.

```
install.packages("devtools") # Run this line if the 'devtools' package isn't installed already.  
devtools::install_github(repo="OuhscBbmc/codified")
```



William Beasley

HTML Output from *codified*

Racial Categories	Ethnic Categories											
	Not Hispanic or Latino			Hispanic or Latino			Unknown/Not Reported Ethnicity					
	Female		Male	Unknown/ Not Reported	Female		Male	Unknown/ Not Reported	Female		Male	Unknown/ Not Reported
	F	M			F	M			F	M		
American Indian/Alaska Native	2	4		0	0	0		0	0	0		
Asian	0	0		0	0	0		0	0	0		
Native Hawaiian or Other Pacific Islander	0	0		0	0	0		0	0	0		
Black or African American	0	0		0	0	0		0	0	0		
White	0	2		0	0	0		0	0	0		
More than One Race	0	0		0	0	0		0	0	0		
Unknown or Not Reported	0	0		0	0	0		0	0	0		

Now with a packagedown website

codified 0.0.1.9001 [Home](#) Reference Articles ▾ Changelog [?](#)

<https://ouhscbbmc.github.io/codified/articles/nih-enrollment-html.html>

NIH Enrollment Tables in HTML

Will Beasley and Peter Higgins

2018-09-06

Source: vignettes/nih-enrollment-html.Rmd

Install codified Package

First, install the `codified` package if necessary, and then [load it into memory](#).

```
if( !requireNamespace("codified", quietly=T) )
  devtools::install_github(repo= "OuhscBbmc/codified")

library(codified)
```

Contents

- [Install codified Package](#)
- [Local Data Source](#)
- [Establish Datasets](#)
- [Apply Map to Observed Dataset](#)
- [Conform to NIH Cosmetics](#)
- [REDCap Data Source](#)
- [Establish Datasets](#)
- [Conform to NIH Cosmetics](#)
- [Collapsing Levels](#)



William Beasley

Thanks To...

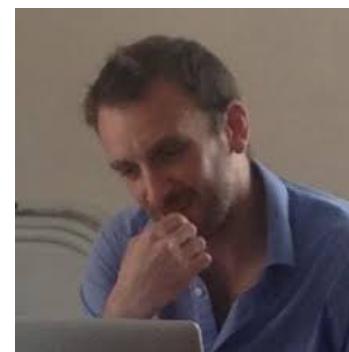
- Use of ***keyring*** package to securely use tokens
- Use of ***RedCapR*** package to extract data via API
- Use of `dplyr::mutate(case_when)` for wrangling
- Use of ***janitor::tabyl*** for creating a 3D counts table
- Use of ***tidyverse::complete*** to fill in all combinations of factors (even ones that did not occur in the data)
- Lots of ***tidyverse*** (unite, separate, and spread)
- Make the table pretty with ***flextable, kableExtra***
- Save directly to Word and Powerpoint with ***officer***



William Beasley



Sam Firke



David Gohel



Questions?

Thanks for your interest!

Convert to a table of counts

```
ibd_table <- ibd %>%  
  tabyl(race, sex2, ethnic_cat) %>% # creates list of 3 tables  
  reduce(left_join, by = "race") # purrr reduces to one table  
  # 20 rows have non-zero counts
```

	race	Female.x	Male.x	Female.y	Male.y	Female	Male
1	Asian	0	1	3	5	0	0
2	Black or African-American	0	0	1	1	0	0
3	More Than One Race	1	0	1	1	0	0
4	Unknown or Not Reported	0	0	0	1	2	1
5	White	5	1	43	46	1	3

Make a Table of all possible combinations of categories, with all counts = 0

```
# make a list of four vectors of length 7
l1 <- list(race = c("White", "Black or African-American", "Asian",
  "Native Hawaiian or Other Pacific Islander", "American Indian or Alaska
  Native", "More Than One Race", "Unknown or Not Reported"),
  sex = c("male", "female", "Unknown or Not Reported Sex", "male", "female",
  "male", "female"),
  ethnicity = c("Hispanic", "Not", "Unknown", "Hispanic", "Not", "Hispanic",
  "Not"),
  count = rep(0,7))
```

> as_tibble(l1)				
	# A tibble: 7 x 4			
	race	sex	ethnicity	count
	<chr>	<chr>	<chr>	<dbl>
1	White	male	Hispanic	0
2	Black or African-American	female	Not	0
3	Asian	Unknown or Not Re...	Unknown	0
4	Native Hawaiian or Other P...	male	Hispanic	0
5	American Indian or Alaska...	female	Not	0
6	More Than one Race	male	Hispanic	0
7	Unknown or Not Reported	female	Not	0

Make a Table of all possible combinations of categories, with all counts = 0

```
empty_table <- as_tibble(1) %>%  
  tidyverse::complete(race,  
    nesting(sex), nesting(ethnicity), fill=list(count = 0))
```

# A tibble: 63 x 4			
	race	sex	ethnicity count
	<chr>	<chr>	<chr> <dbl>
1	American Indian or Alaska Native	female	Hispanic 0
2	American Indian or Alaska Native	female	Not 0
3	American Indian or Alaska Native	female	Unknown 0
4	American Indian or Alaska Native	male	Hispanic 0
5	American Indian or Alaska Native	male	Not 0
6	American Indian or Alaska Native	male	Unknown 0
7	American Indian or Alaska Native	Unknown or ...	Hispanic 0
8	American Indian or Alaska Native	Unknown or ...	Not 0
9	American Indian or Alaska Native	Unknown or ...	Unknown 0

Now filter 2 tables

- Gather and filter actual counts to only have non-zero rows

```
ibd_table2 <- gather(ibd_table, key= sex.eth, value = count, -race) %>%  
separate(sex.eth, into = c('sex', 'ethnicity')) %>%  
filter(count != 0) # 20 non-zero rows
```

- Anti-join with empty table to get only the needed zero rows

```
complement <- anti_join(empty_table, ibd_table2,  
by = c('race', 'sex', 'ethnicity'))  
# complement is 43 rows with zeros
```

Now make full 63 row table

```
full_table <- bind_rows(ibd_table2, complement)
```

DEMO – create table of counts

tabyl x 2, tabyl x 3, tabyl x 3 with reduce

- Now it is tidy and complete.
- Now to make it un-tidy for standard formatting
 - Sex is nested within ethnicity in the NIH table
 - Still need to add row totals and column totals

Wrangling nested sex, ethnicity

- Unite to nest sex within ethnicity as eth.sex
- Then spread

```
ibd_table <- full_table %>%  
  unite(col = "eth.sex", c('ethnicity', "sex"), sep=".") %>%  
# three cols - race, eth.sex, count  
  spread(key = eth.sex, value = count)  
# now spread to 10 cols
```

DEMO wrangle nested sex within ethnicity

Adding Margin totals

```
# convert race col to rownames to make numbers into a matrix
m <- as.matrix(ibd_table[ , -1]) # removes col1 (race) in m
rownames(m) <- ibd_table$race # saves race in rownames

ibd_table2 <- addmargins(m, FUN=c(Total=sum), quiet = T)

ibd_table <- rownames_to_column(as.data.frame(ibd_table2),
"Racial Categories")
# puts race back into dataframe from rownames
```

DEMO add margin totals

Now to make nice tables

- ***Flextable*** is a package by David Gohel (officer, reporters)
 - makes nicely formatted tables for Word and Powerpoint
 - Very well documented, multiple vignettes
 - Functionality similar to the kableExtra package for HTML
- Works well with Rmd (HTML), .docx, .pptx
 - Not so much with LaTex, PDF if that is your thing
 - ***DEMO make flextable
show myft after add each header, before mergeh, mergev, final***
 - https://github.com/higgi13425/nih_enrollment_table

Output to Word, Powerpoint

- ***officer*** package

```
doc <- read_docx()  
doc <- body_add_flextabLe(doc, value = myft)  
print(doc, target = "/path/file.docx")  
  
ppt <- read_pptx()  
ppt <- add_slide(ppt, layout = "Title and Content",  
                  master = "Office Theme")  
ppt <- ph_with_flextabLe(ppt, value = myft, type= "body")  
print(ppt, target = "/path/file.docx")
```