HUDK 4050:
CORE METHODS
IN EDM

9/26/17 10:55 AM

In the news

SCIENTIFIC AMERICAN

Intelligence and the DNA Revolution

Scientists identify 22 genes associated with intelligence

The False Prophecy of Hyperconnection

FOREIGN AFFAIRS

THE CHRONICLE OF HIGHER EDUCATION

The Mismatch Between High Tech and Higher Ed

TIMESDAILY

Edtech is the next fintech

Forget what you've been told about edtech

Kids Count: Poverty rates are up; hunger stats have stabilized

The Secret Sauce to Choosing Edtech? Find Tools By Fit, Not Feature

EdSurge

CAMPUS TECHNOLOGY

How to Protect Education Data When No Systems Are Secure

Education Data Breaches Double in First Half of 2017

U.S.News

3 Lessons Learned From Education Technology Research

EdTech Focus On K-12

CDE CO COLORADO Department of Education

Solving Real-World Problems Is Key to Ed Tech Success

Do the Benefits Outweigh the Disadvantages of Online SATs?

McAfee Finds Teenagers Use Tech to Cheat in School

Visualize Colorado's Education Data By District

Thurrott

Microsoft 365 Evolves, Comes to Education

THE VERGE

Two edtech champions to join White House offices as fellows

edscoop

AI trained on Yelp data writes fake restaurant reviews 'indistinguishable' from real deal

Events

Event	Date	Time	Location	URL
Building Better Digital Citizens: Obama Foundation	September 28	6:30pm	Secret	https://www.meetup.com/betanyc/events/243584327/?https=on&_af=event&_af_eid=243584327
Regulating Online Conversation in the Platform Era	October 10	7:00pm	New School	http://events.newschool.edu/event/free_speech_hate_speech_regulating_online_conversation_in_the_platform_era#.WcK3TEyZOL5
Deep Learning	October 10	all day	Montreal, Canada	https://www.re-work.co/events/deep-learning-summit-montreal-canada-track1-2017
AWS EdStart Pitch	October 4	6:00pm	350 W Broadway	https://pages.awscloud.com/NAMER_STARTUP_loft-ny-EdStartPitchDay_October_2017_Registration.html
Empower The User Breakfast	September 28	8:00am	Irish Consulate	https://www.meetup.com/New-York-Risk-Meetup/?gj=ej1b
Intro to Shiny Webinar	September 27	11:00am	Online	https://pages.rstudio.net/September27_Registration.html?mkt_tok=eyJpIjoiTXpZd01ESXpabVE1TkdKayIsInQiOiJxN0VidHluWXg4NXl1cWdWWWQ3RlV5MWNVbGp5TmR5ZjhGMkV3SWFaWVhQZ1ZYVUk0a1d4anVyK2RKSVViT2hnMEhNNDFXXC9tazJGZmU2RkVlODNyMW8xcVowZ1BzV2swSndld2gzOTdkQWdYc2ZNVWQxZXVRYVZxcXhGUDl6RFUifQ%3D%3D
The People's Disruption: Platform Co-Ops for Global Challenges	November 10/11	all day	The New School	https://platform.coop/2017

Opportunities

Data Intern for Hope Education Research Solutions
Leslie Ponciano
Send CV to lesliеponciano@gmail.com

Data & Society
Communications Production Manager
“Asana master to manage the production process for our research products”

Today

- R Markdown
- Gather & Spread
- Swirl data manipulation activity

R Markdown

http://rmarkdown.rstudio.com/gallery.html

(Charles, this is to remind you to demonstrate what the interface looks like)

Data Wrangling
with dplyr and tidyr
Cheat Sheet

RStudio

Tidy Data - A foundation for wrangling in R

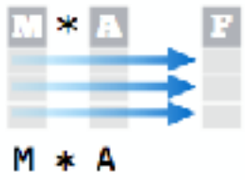
In a tidy data set:

Each **variable** is saved in its own **column**

&

Each **observation** is saved in its own **row**

Tidy data complements R's **vectorized operations**. R will automatically preserve observations as you manipulate variables. No other format works as intuitively with R.



Syntax - Helpful conventions for wrangling

dplyr::tbl_df(iris)

Converts data to tbl class. tbl's are easier to examine than data frames. R displays only the data that fits onscreen:

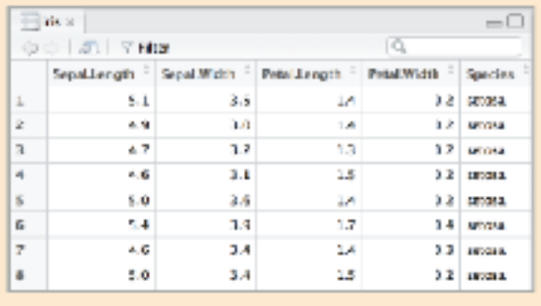
```
Source: local data frame [150 x 5]

   Sepal.Length Sepal.Width Petal.Length
1           5.1         3.5          1.4
2           4.9         3.0          1.4
3           4.7         3.2          1.3
4           4.6         3.1          1.5
5           5.0         3.6          1.4
..          ...         ...          ...
Variables not shown: Petal.Width (dbl),
  Species (fctr)
```

dplyr::glimpse(iris)

Information dense summary of tbl data.

utils::View(iris)

View data set in spreadsheet-like display (note capital V).



dplyr::%>%

Passes object on left hand side as first argument (or . argument) of function on righthand side.

```
x %>% f(y)  is the same as  f(x, y)
y %>% f(x, ., z)  is the same as  f(x, y, z)
```

"Piping" with %>% makes code more readable, e.g.

```
iris %>%
  group_by(Species) %>%
  summarise(avg = mean(Sepal.Width)) %>%
  arrange(avg)
```

Reshaping Data - Change the layout of a data set

tidyr::gather(cases, "year", "n", 2:4)

Gather columns into rows.

tidyr::spread(pollution, size, amount)

Spread rows into columns.

tidyr::separate(storms, date, c("y", "m", "d"))

Separate one column into several.

tidyr::unite(data, col, ..., sep)

Unite several columns into one.

dplyr::data_frame(a = 1:3, b = 4:6)

Combine vectors into data frame (optimized).

dplyr::arrange(mtcars, mpg)

Order rows by values of a column (low to high).

dplyr::arrange(mtcars, desc(mpg))

Order rows by values of a column (high to low).

dplyr::rename(tb, y = year)

Rename the columns of a data frame.

Subset Observations (Rows)

dplyr::filter(iris, Sepal.Length > 7)

Extract rows that meet logical criteria.

dplyr::distinct(iris)

Remove duplicate rows.

dplyr::sample_frac(iris, 0.5, replace = TRUE)

Randomly select fraction of rows.

dplyr::sample_n(iris, 10, replace = TRUE)

Randomly select n rows.

dplyr::slice(iris, 10:15)

Select rows by position.

dplyr::top_n(storms, 2, date)

Select and order top n entries (by group if grouped data).

Logic in R - ?Comparison, ?base::Logic

<	Less than	!=	Not equal to
>	Greater than	%in%	Group membership
==	Equal to	is.na	Is NA
<=	Less than or equal to	!is.na	Is not NA
>=	Greater than or equal to	&, |, !, xor, any, all	Boolean operators

Subset Variables (Columns)

dplyr::select(iris, Sepal.Width, Petal.Length, Species)

Select columns by name or helper function.

Helper functions for select - ?select

select(iris, contains("."))
Select columns whose name contains a character string.

select(iris, ends_with("Length"))
Select columns whose name ends with a character string.

select(iris, everything())
Select every column.

select(iris, matches(".t."))
Select columns whose name matches a regular expression.

select(iris, num_range("x", 1:5))
Select columns named x1, x2, x3, x4, x5.

select(iris, one_of(c("Species", "Genus")))
Select columns whose names are in a group of names.

select(iris, starts_with("Sepal"))
Select columns whose name starts with a character string.

select(iris, Sepal.Length:Petal.Width)
Select all columns between Sepal.Length and Petal.Width (inclusive).

select(iris, -Species)
Select all columns except Species.

devtools::install_github("rstudio/EDAWR") for data sets

Learn more with browseVignettes(package = c("dplyr", "tidyr")) • dplyr 0.4.0• tidyr 0.2.0 • Updated: 1/15

Reshape

- Similar to generating pivot tables
- Long format ⟷ Wide format

Student	Quiz 1	Quiz 2	Quiz 3
Francis	10	10	11
Alex	14	15	18
Kaji	11	17	14
Miriam	8	10	8

Spread ←

→ **Gather**

Student	Quiz	Date
Francis	10	Quiz 1
Francis	10	Quiz 2
Francis	11	Quiz 3
Alex	14	Quiz 1

tidyr::gather()

Specify:

- dataframe
- key: the variable that the reshape will be based on
- value: new variable names that will generated for new data frame
- gather_cols: the variables that are reshaped to accommodate the new structure
    - ✳ Can also identify key using “-“ and all other columns will be gathered

tidyr::gather()

Student	Quiz 1	Quiz 2	Quiz 3
Francis	10	10	11
Alex	14	15	18
Kaji	11	17	14
Miriam	8	10	8

Student	Quiz	Date
Francis	10	Quiz 1
Francis	10	Quiz 2
Francis	11	Quiz 3
Alex	14	Quiz 1

key

Student
Francis
Francis
Alex

Quiz 1	Quiz 2	Quiz 3	10	11	14	15	18
10	10	11	Quiz 1	Quiz 3	NA	NA	NA
10	10	11	Quiz1	NA	NA	NA	NA
14	15	18	NA	NA	Quiz 1	Quiz 2	Quiz 3

tidyr::spread()

Specify:

- dataframe
- key: the variable that the reshape will be based on
- value: column whose values will populate the cells

tidyr::spread()

Key **Value**

Student	Quiz	Date
Francis	10	Quiz 1
Francis	10	Quiz 2
Francis	11	Quiz 3
Alex	14	Quiz 1

Student	Quiz 1	Quiz 2	Quiz 3
Francis	10	10	11
Alex	14	15	18
Kaji	11	17	14
Miriam	8	10	8

Class Activity 1

http://bit.ly/2jZOMyH