

# Workshop - Hackathon!

Andrew Stewart

Andrew.Stewart@manchester.ac.uk



@ajstewart\_lang

# Today

- In groups you are going to work on hacking a large dataset - the World Happiness Report Data. I want you to tidy and wrangle as necessary, visualise components, run appropriate statistical tests, and the put all this in a markdown document (rendered as a .html file)
- In your group, decide what activities you're going to do with the dataset.

# Why?

- Real world data is messy - I want you to experience making sense of that.
- Real world research involves collaborating with others - again, this is good practice for managing that.
- One of the best ways to learn new coding/data/statistical tricks/techniques is to see how others do things!
- The Hackathon! assignment (which you will each do individually) will be much easier once you've gone through an actual Hackathon.

# For the assignment

- You need to do a Hackathon individually on a new dataset (i.e., not the one you're looking at as a group today).
- The dataset could be an open dataset from an area you research (or are interested in) - perhaps it was published with a paper.
- Or you could use an open dataset on any topic that interests you.
- Whatever set you choose, I want to see evidence of data wrangling and tidying, visualisation, and modelling - with a summary of what meaning you have extracted from the data (and any caveats about the interpretation that you think are worth raising).

# Good places to start looking for open data sets

- In your research area, there are likely to be large datasets that have already been published - or you could check out and use data from...

- The Google dataset search toolbox:

<https://toolbox.google.com/datasetsearch>

- The Tidy Tuesday datasets:

<https://github.com/rfordatascience/tidytuesday>

- The gapminder datasets:

<https://www.gapminder.org/data/>

- The Kaggle datasets:

<https://www.kaggle.com/datasets>

- Or any other source you might want to use! The data don't have to be psychological in nature.

# A dataset

Using the Google dataset search, I looked for the World Happiness data - to download it, I had to create a free account (not always required):

<https://data.world/laurel/world-happiness-report-data>

data.world/laurel/world-happiness-report-data

Overview Contributors Discussion Activity

2 Coverage, Summary Statistics and Regression Tables

Source: Helliwell, J., Layard, R., & Sachs, J. (2017). World Happiness Report 2017, New York: Sustainable Development Solutions Network.

DATA DICTIONARY View all definitions for the 2 files and the 59 columns in this dataset. View

2 files

online-data-chapter-2-whr-2017.xlsx

data\_behind\_table\_2\_1\_whr\_2017 figure\_2\_1\_whr2017 figure2\_2\_whr\_2017 figure2\_3\_whr\_2017

	wp5_country	country	year	# life_ladder	# log_gdp_per_capita	# social_support
1	Afghanistan	Afghanistan	2008	3.7236	7.1971	0.4507
2	Afghanistan	Afghanistan	2009	4.4018	7.3627	0.5523
3	Afghanistan	Afghanistan	2010	4.7584	7.4163	0.5391
4	Afghanistan	Afghanistan	2011	3.8317	7.4458	0.5211
5	Afghanistan	Afghanistan	2012	3.7829	7.5492	0.5206

Showing 1-5 of 1,420 rows, 27 columns See all

Switch to column overview

RECENT COMMENTS

@kieroneil · 9 months ago  
Nevermind, I see that Life Ladder is the Happiness Score

@kieroneil · 9 months ago  
Is the 2016 data in the first sheet used to determine the 2017 scores i

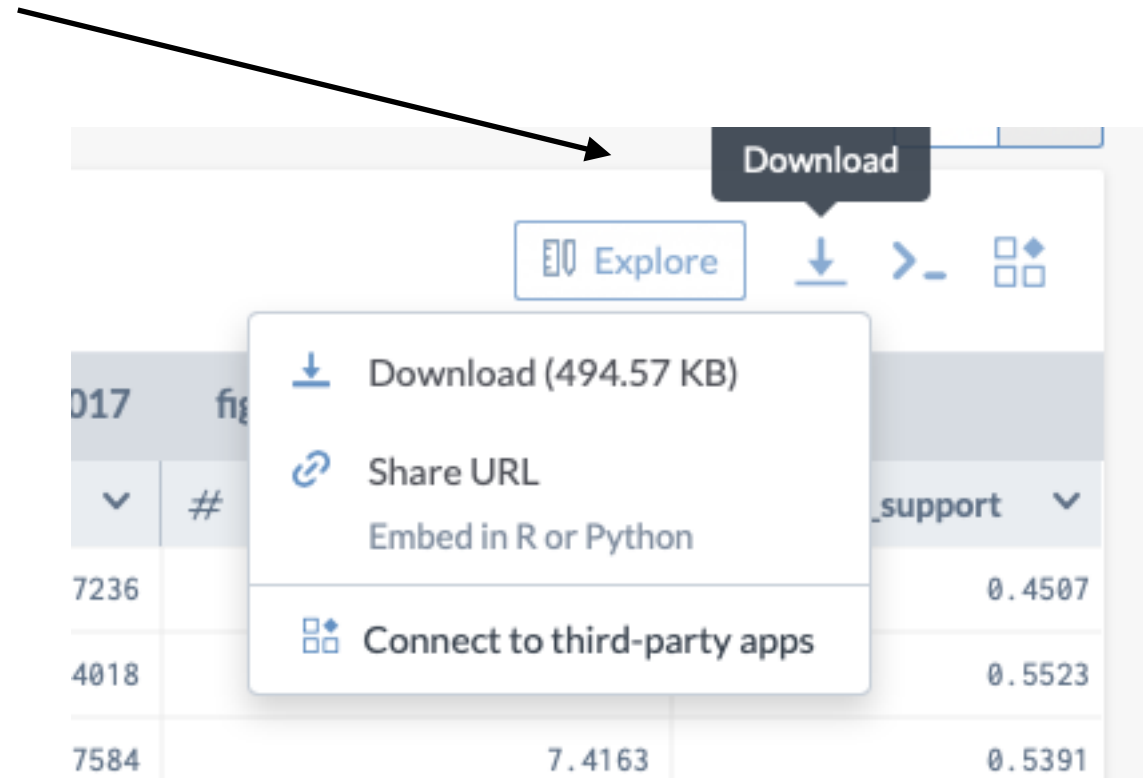
@quanticdata · last year  
https://public.tableau.com/profile/st

@noahbohnson · last year  
Zoomed in a bit on Eurasia/Africa















@noahbohnson · last year  
Average Life Ladder by Country from Tableau

View comments

Click the download icon and you can either download the file, or copy the link to open in R:



Click on the Data Dictionary button for an explanation as to what each column represents.

  Data dictionary 		
Data dictionary		
data_behind_table_2_1_whr_2017		
 wp5_country 	string	WP5 is GWP's coding of countries, including some sub-country territories such as Hong Kong.
 country 	string	Country
 year 	year	Year
# life_ladder 	decimal	Happiness score or subjective well-being (variable name ladder ).
# log_gdp_per_capita 	decimal	Statistics of GDP per capita (variable name gdp) in purchasing power parity (PPP) at constant 2011 int. dollar prices
# social_support 	decimal	Social support (or having someone to count on in times of trouble) is the national avg of the binary responses (0 or 1)
# healthy_life_expectancy_at_birth 	decimal	The time series of healthy life expectancy at birth calculated by WHO, WDI, and other published stats
# freedom_to_make_life_choices 	decimal	National avg responses to "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"



If you copy the link, then you can read an Excel file from a website into R like:

```
library(tidyverse)
library(readxl)

url1 <- "https://query.data.world/s/tw3oaknxjlqods27xzzbpa3do4rmfr"
p1f <- tempfile()
download.file(url1, p1f, mode="wb")
happy_data <- read_excel(path = p1f)
```

Just replace the url1 link with the one that you've copied via (in this case) the Share URL option...

First get into a group. Now, here are some tasks you need to do with the Happiness survey dataset:

1. tidy and wrangle as needed
2. visualise - there are lots of variables to visualise
3. model - there are several factors and multiple DVs you could look at
4. as you go, email me your script chunks as you complete them

Each time you complete a sub-task task in your group, email me your R code, and move onto another task.

Feel free to stop for a snack, lunch whenever you want - you can eat in this room if needs be.

Towards the end of the day (at around 1515) I will knit a markdown file as an .html file that presents the classes work today. Give you're doing this in .html it would be nice to see an animation or two!

- All activity will be co-ordinated via this site:

[https://hackmd.io/\\_2008xkVSiGtELYQXoEqHg](https://hackmd.io/_2008xkVSiGtELYQXoEqHg)

- Get into your groups.
- Load the Happiness data file into R.
- Work as a group from now until 1500 - at 1515 I'll knit all your code together.
- Each time you complete a sub-task - email me your R script.

hackmd.io/\_2008xkVSiGtELYQXoEqHg?both

Google ScholarScopusjobs.ac.ukBBC NewsChester WeatherThe GrauniadThe IndependentGoogle MapsChester Weather StationGitHub

HackMD

+

?

New! New Feature: account merge!

...

Sign in

1 ONLINE

B I S H

</> “ ≡ ≡ ✓

🔗 🖼️ 📄 - 💬

1# Psychology MRes Hackathon 4/4/19

2

3

4Download the Happiness survey data (or use a different dataset of your own choosing if you like) - note, for the assignment you CANNOT use the Happiness survey data.

5

6The data dictionary can be found here:

7

8<https://data.world/laurel/world-happiness-report-data/workspace/data-dictionary>

9

10You can paste the following code into your R script to load the data and map it into the tibble happy\_data.

11

12> library(tidyverse)

13> library(readxl)

14>

15> url1 <-  
"https://query.data.world/s/tw3oaknxjlqods27xzzbpa3do4rmfr"

16> p1f <- tempfile()

17> download.file(url1, p1f, mode = "wb")

18> happy\_data <- read\_excel(path = p1f)

19

20First get into a group. Now, here are some tasks you need to do with the Happiness survey dataset:

Line 32, Columns 59 — 38 Lines

✓

○

Spaces: 4

SUBLIME

🔧

Length: 1599

⌚ CHANGED A FEW SECONDS AGO

## Psychology MRes Hackathon 4/4/19

Download the Happiness survey data (or use a different dataset of your own choosing if you like) - note, for the assignment you CANNOT use the Happiness survey data.

The data dictionary can be found here:

<https://data.world/laurel/world-happiness-report-data/workspace/data-dictionary>

You can paste the following code into your R script to load the data and map it into the tibble happy\_data.

```
library(tidyverse)
library(readxl)

url1 <- "https://query.data.world/s/tw3oaknxjlqods27xzzbpa3do4rmfr"
p1f <- tempfile()
download.file(url1, p1f, mode = "wb")
happy_data <- read_excel(path = p1f)
```

First get into a group. Now, here are some tasks you need to do with the Happiness survey dataset:

1. tidy and wrangle as needed
2. visualise - there are lots of variables to visualise
3. model - there are several factors and multiple DVs you could look at

# Handy hints

- Remember I will bring all your code together at the end so if you are changing variable names make sure that's captured in the code you email to me.
- Any questions?