Workshop - Hackathon!

Andrew Stewart

Andrew.Stewart@manchester.ac.uk

 @ajstewart_lang

Today

- In one large group 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)

- You can split into smaller groups if you like (or work individually) on a specific aspect of the above - you decide.

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/staistical 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:

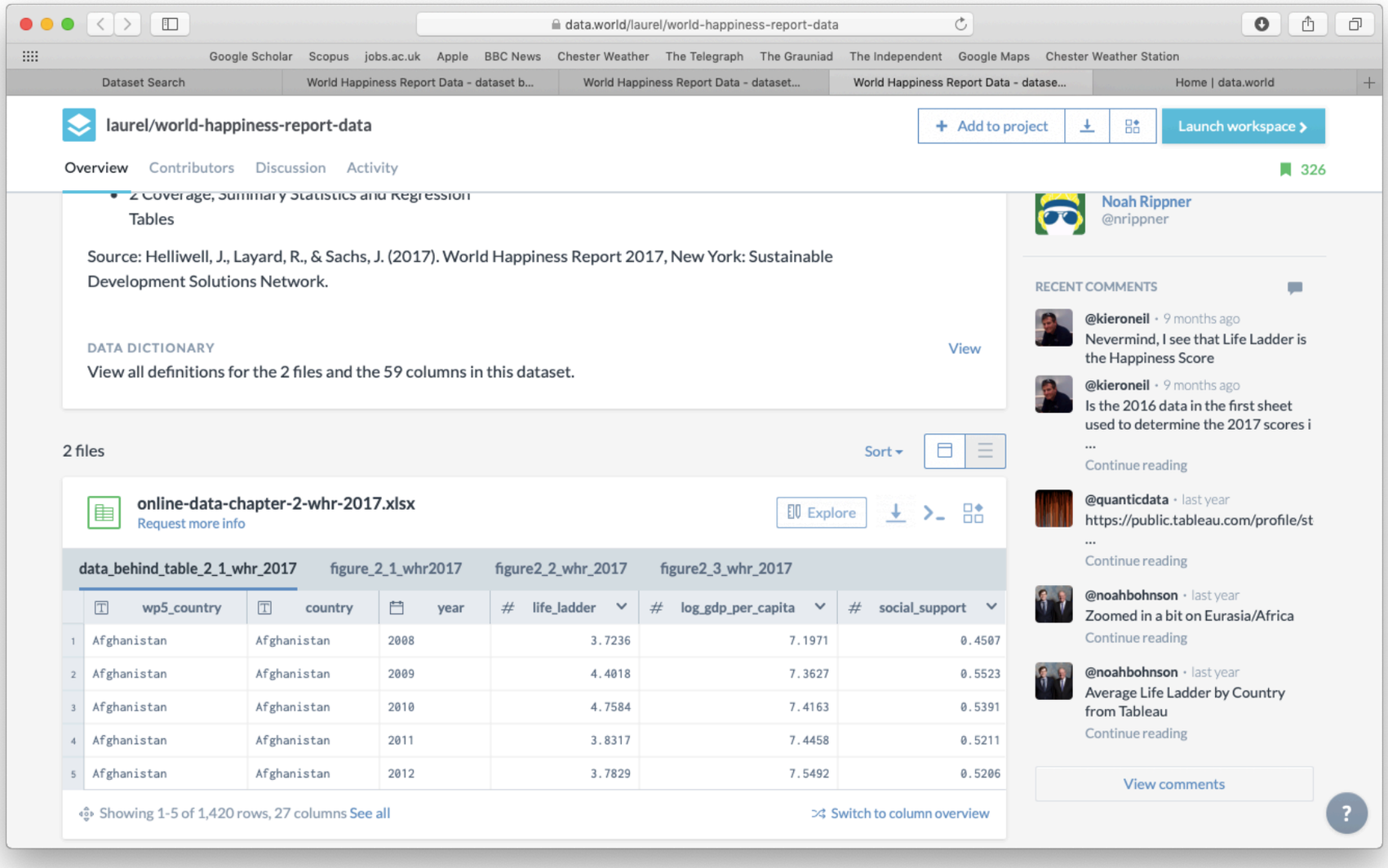
  https://www.kaggle.com/datasets

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

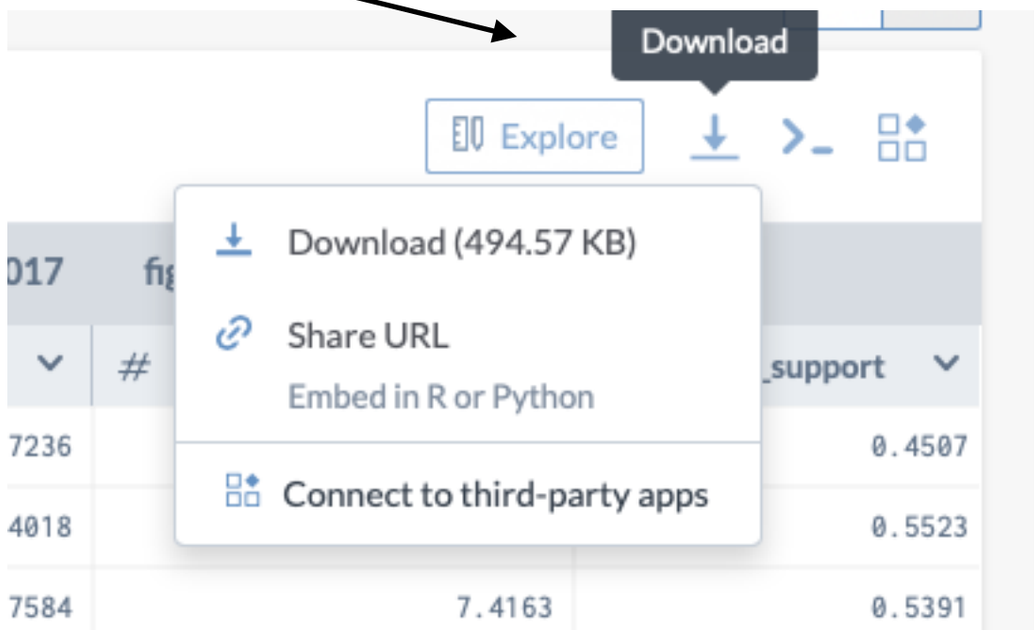
A dataset

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



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



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…

Here are some tasks you need to do:

1. tidy and wrangle
2. visualise - there are lots of variables to visualise
3. model - there are several factors and DVs you could look at
4. report back - bring together the above to create a markdown file which brings all of the above together

Each group should put their group or individual names alongside what activities you are going to engage with. I suspect you'll mainly be working on visualising or modelling. Make it clear which variables you will be working with when you write your information below.

Towards the end of the day (at around 1500) 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

- Get into your groups (or decide to work individually).

- Download the Happiness data file (or find a different one if you like).

- Decide what each of you (individually or in pairs) is going to do (e.g., mainly developing visualisation or mainly doing stats modelling) and write that down on the hackmd.io page so people don't end up all doing the same thing.

- Work from now until 1500 - at 1500 I'll knit all your code together.

- Each time you complete a task - email me your R script.

hackmd.io/_2008xkVSIGtELYQXoEqHg?both

HackMD

New! New Feature: account merge!

Sign in

1 ONLINE

```
# 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.

> library(tidyverse)
> library(readxl)
>
> url1 <-
"https://query.data.world/s/tw3oaknxjlqods27xzzbpa3d
o4rmfr"
> p1f <- tempfile()
> download.file(url1, p1f, mode = "wb")
> happy_data <- read_excel(path = p1f)

Here are some tasks you need to do:
1. tidy and wrangle
2. visualise - there are lots of variables to
visualise
3. model - there are several factors and multiple
DVs you could look at
4. report back - bring together the above to create
a markdown file which brings all of the above
together

Each group should put their group or individual
```

Line 23, Columns 1 — 28 Lines | Spaces: 4 | SUBLIME | Length: 1354

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.

```
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)
```

Here are some tasks you need to do:

1. tidy and wrangle
2. visualise - there are lots of variables to visualise
3. model - there are several factors and multiple DVs you could look at
4. report back - bring together the above to create a markdown file which brings all of the above together

Each group should put their group or individual names alongside what activities you are going to engage with. I suspect you'll mainly be working on visualising or modelling. Make it clear which variables you will be working with when you write your information below.

Towards the end of the day (at around 1500) I will knit a markdown file as

Handy hints

- Co-ordinate activity - do this via making it clear on hackmd.io what it is you are doing - make sure you all update the hackmd.io document as you go along - if you've completed an activity that you've put beside your name, write "COMPLETED".

- 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?