Workshop - Hackathon! Zombie Apocalypse Edition

Andrew Stewart

Andrew.Stewart@manchester.ac.uk

@ajstewart_lang

Workshop	Topic
1	Mixed Models (Andrew)
2	Bayesian Statistics (Johan)
3	Advanced R (Andrew)
4	Matlab (Bo)
5	Hackathon (Andrew)

<u>Assignment hand in:</u>

Mixed Models - Feb 28th

Hackathon - May 22nd

Today

- Today 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 generate a html file via R markdown.

- This will give you practice that will help in the Hackathon! assignment.

Why?

- Real world data is messy - I want you to experience making sense of that.
- One of the best ways to learn new coding/data/statistical tricks/techniques is to see how others do things - we'll share your html documents at the end of the session today.
- 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 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).

- Marks will be awarded for (as usual) good coding, good narrative, clear visualisation, appropriate statistical modelling and interpretation. Extra marks will be awarded if you use packages/functions that we didn't cover in class.

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:

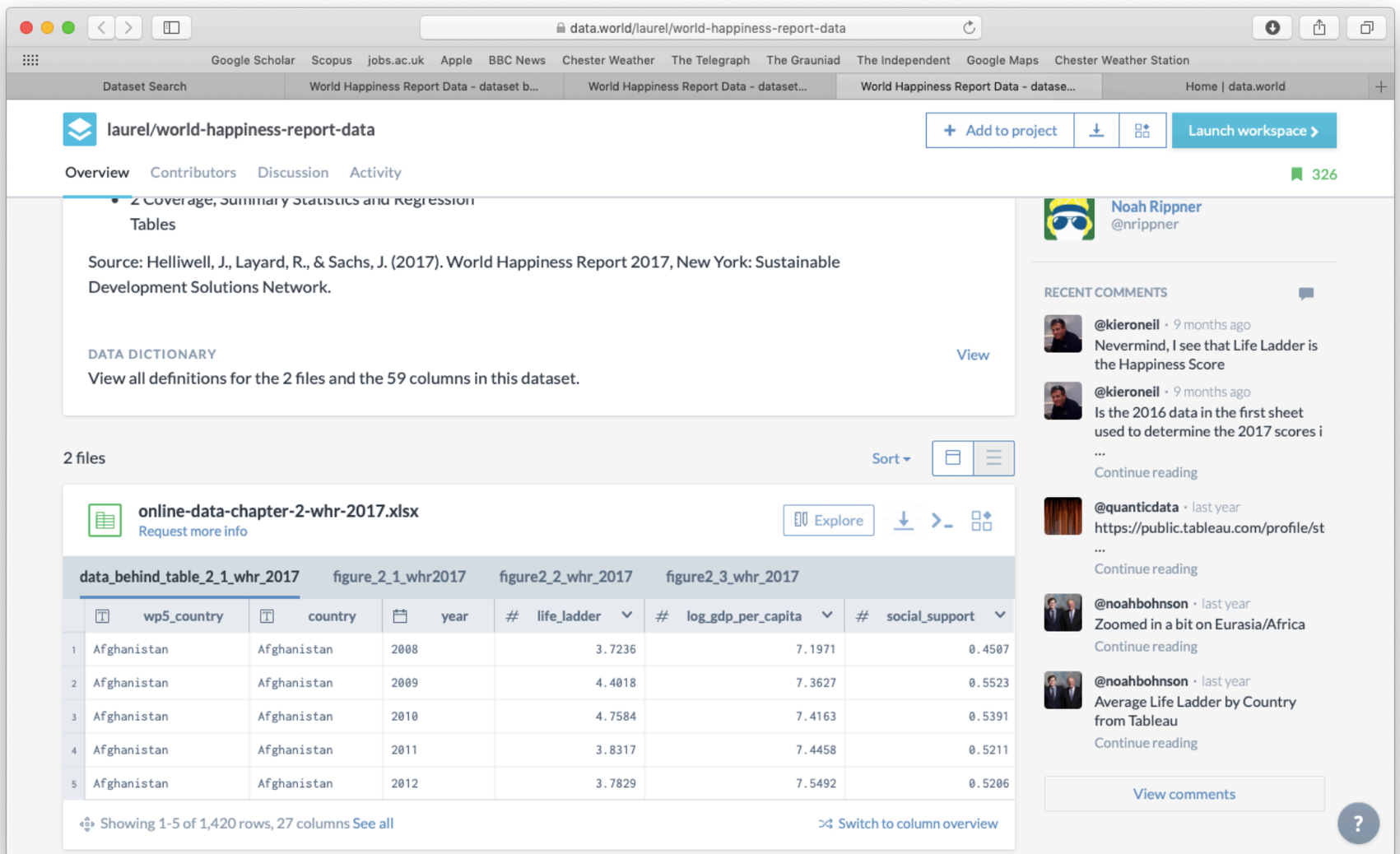
  https://www.kaggle.com/datasets

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

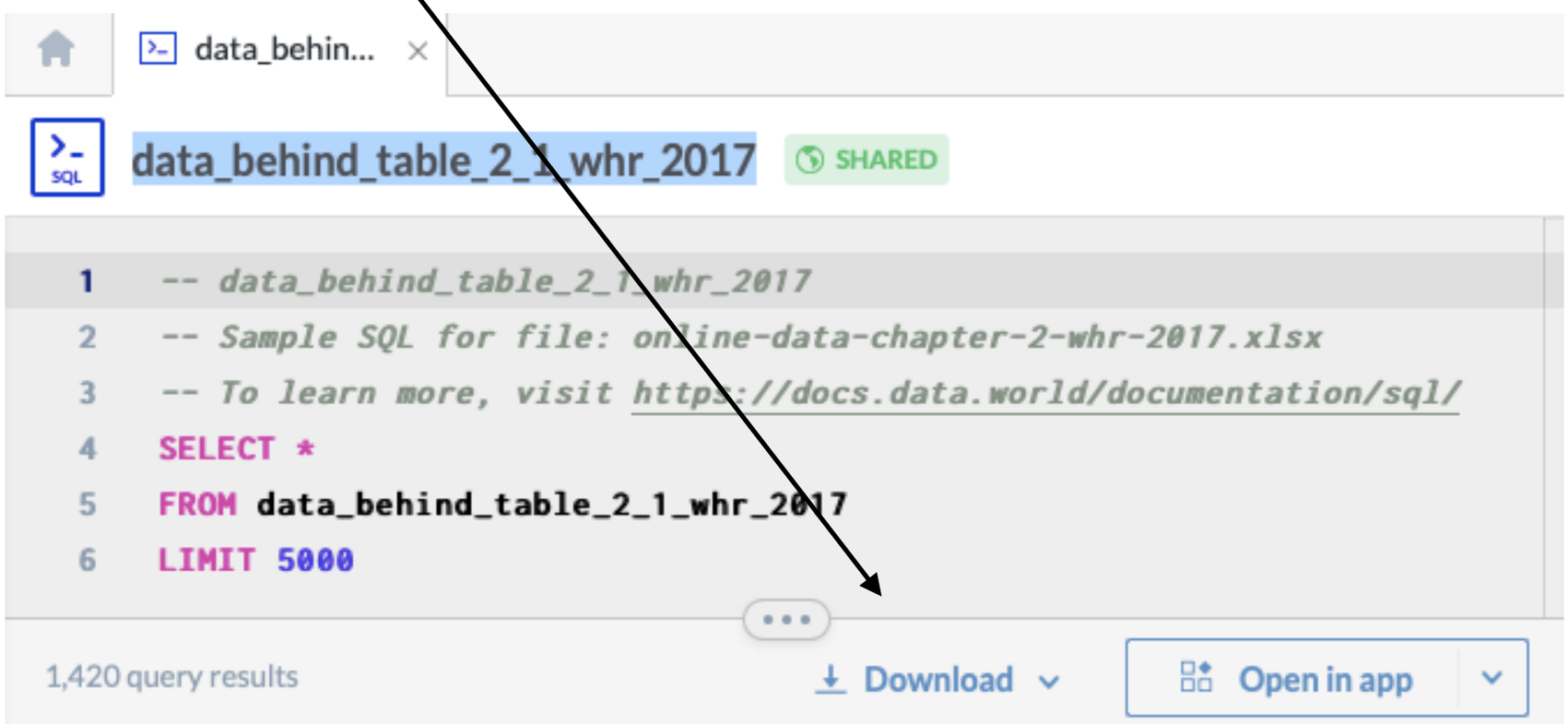
A dataset

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



Open the data file
"data_behind_table_2_1_whr_2017"
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

COLUMN NAME	TYPE	DESCRIPTION
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…

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

You'll probably find it easier/quicker to write your R code in an R Markdown document - you'll need to knit this to html towards the end of the day - probably around 1500/1515.

By **1530 at the latest** email the html file to me - I'll add everyone's contributions to the GitHub repository for this course and share the link with you. Please make sure you have your name at the top of your html document.