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 then put all this in a markdown document.

- 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:

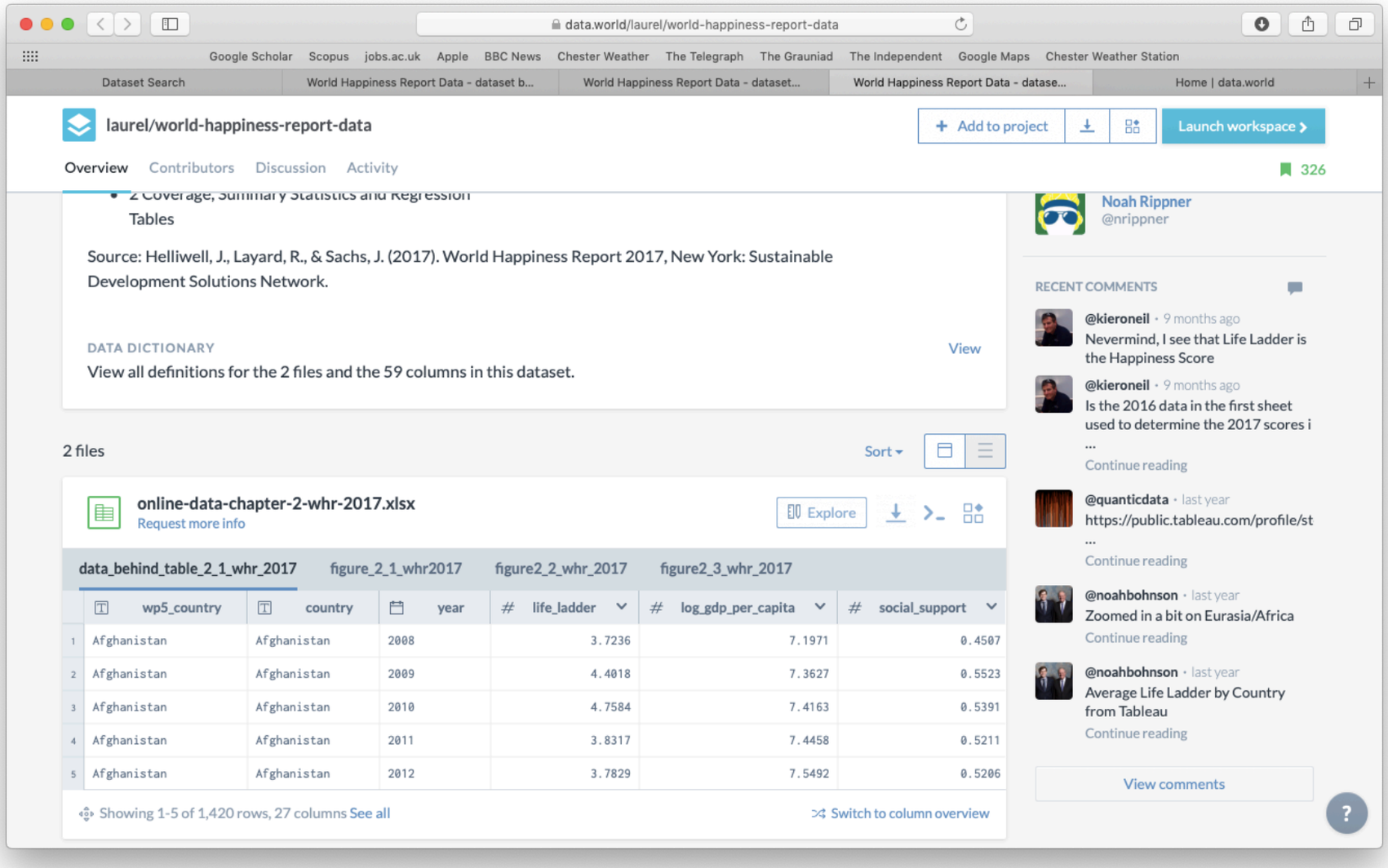
  https://www.kaggle.com/datasets

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

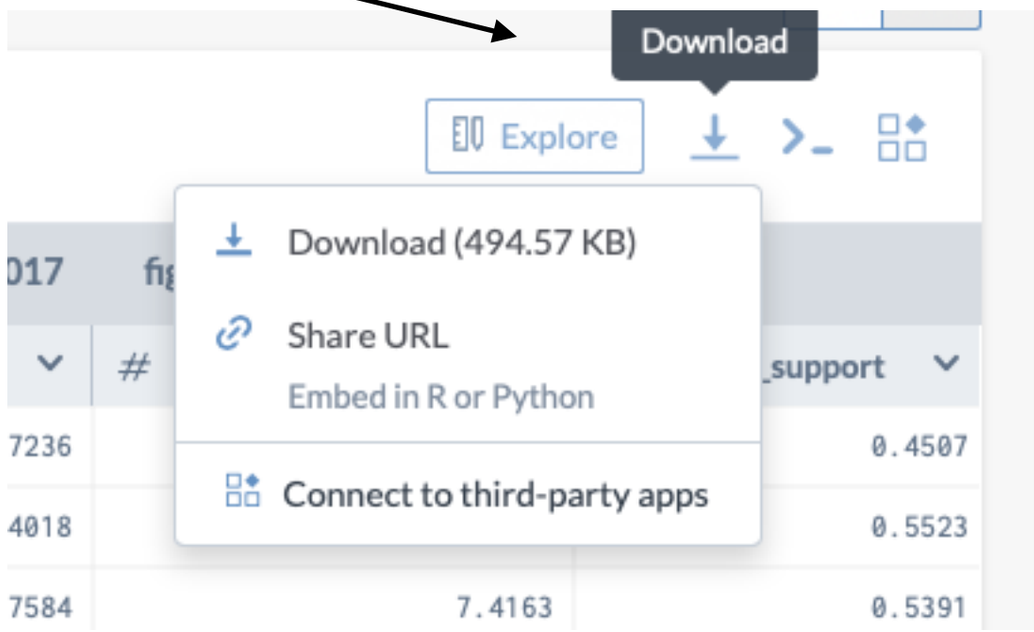
A dataset

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):

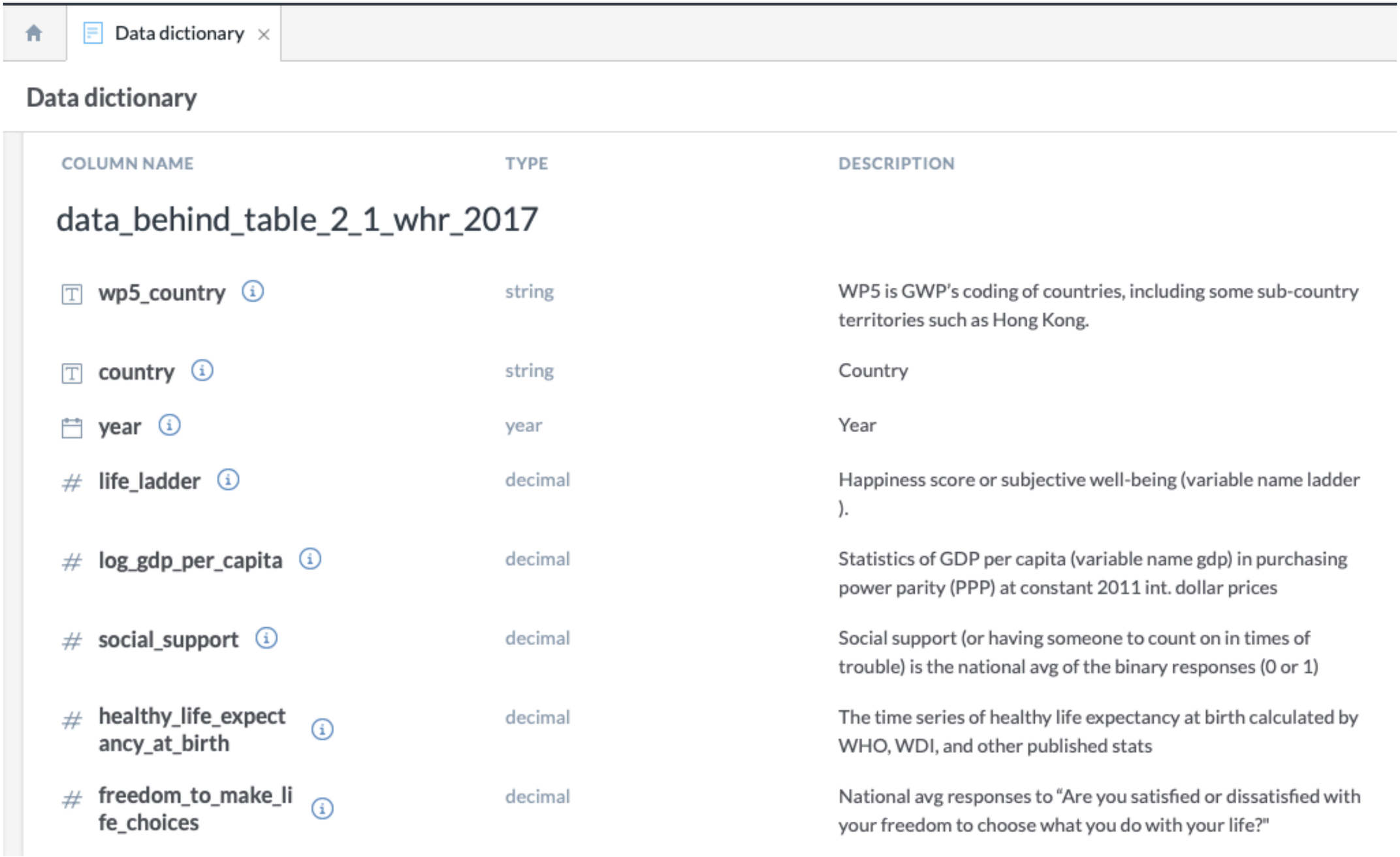
https://data.world/laurel/world-happiness-report-data



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.



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

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

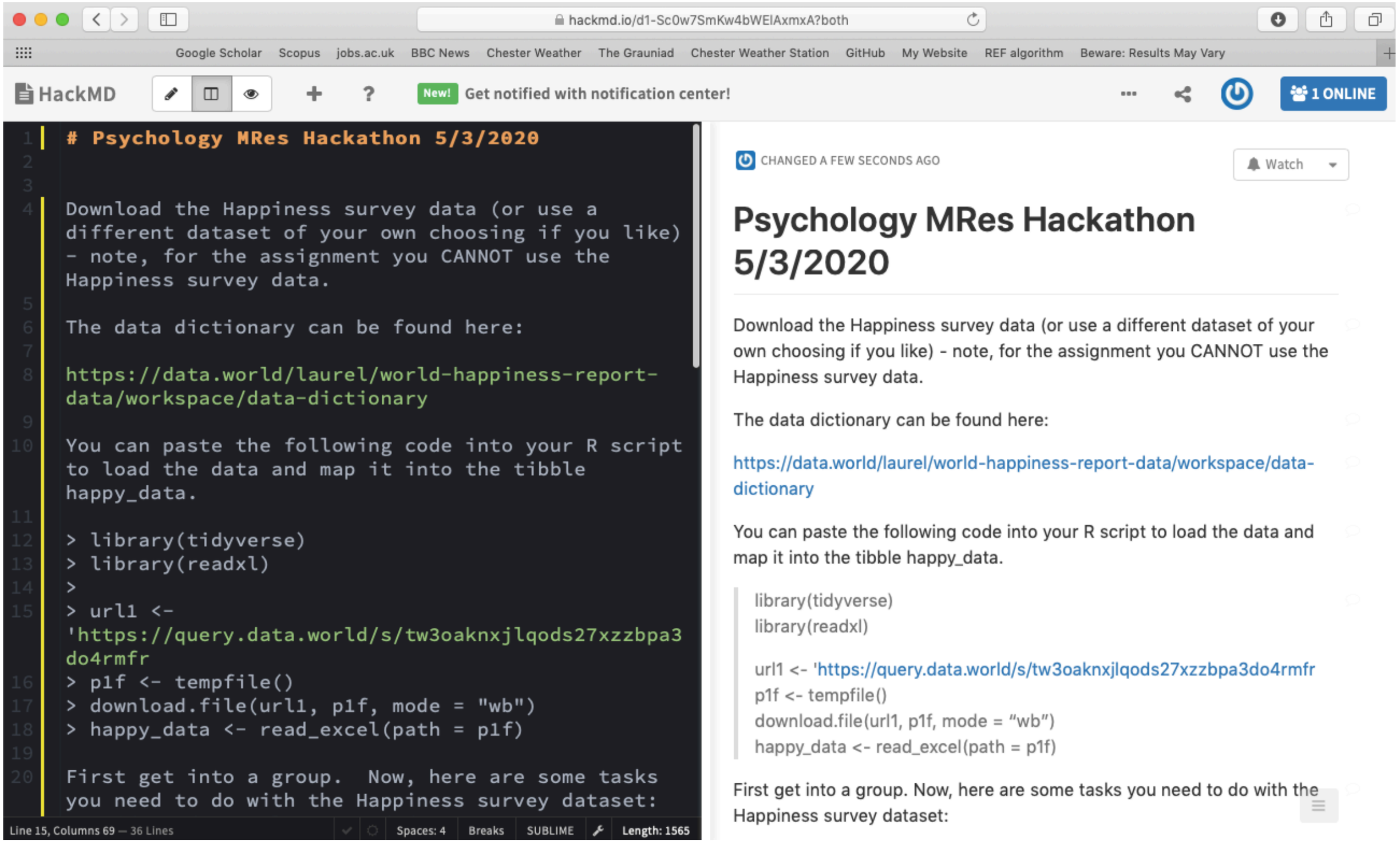
Feel free to stop for a snack, lunch whenever you want - you can eat in this room if needs be.

Given 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/@ajstewartlang/r1L4R87eI/edit`

- Get into your groups.
- Load the Happiness data file into R.
- Work as a group from now until 1530.


hackmd.io/d1-Sc0w7SmKw4bWElAxmxA?both
Google Scholar
Scopus
jobs.ac.uk
BBC News
Chester Weather
The Grauniad
Chester Weather Station
GitHub
My Website
REF algorithm
Beware: Results May Vary
HackMD
New!
Get notified with notification center!
1 ONLINE
Psychology MRes Hackathon 5/3/2020
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:
Line 15, Columns 69 — 36 Lines
Spaces: 4
Breaks
SUBLIME
Length: 1565
CHANGED A FEW SECONDS AGO
Watch
Psychology MRes Hackathon 5/3/2020
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:

- Any questions?