

AGE RANGE

14-16 years

LESSON TYPE

Visual / block-based coding

REQUIREMENTS

- Modern web browser
- Internet connection

# ANALYSING DATA USING BLOCK-BASED PROGRAMMING

When teaching topics from the field data, rich data sets are often necessary. Twitter data, one of the richest data sources on the internet, can be used in school

**F**ostering data-orientated competencies is more and more often considered an important goal for school. Students should learn what can be done with data and – at best – should be able to conduct simple data analyses by themselves. In this lesson plan, we present a simple approach for introducing this topic in school. By analysing data

using the easy-to-start block-based programming language, Snap!, students will gain first insights into how data analyses work and what can be done with data, while also discussing the validity of different analysis approaches.

For this lesson, we selected Twitter as a data source. Although at first glance, Twitter data mainly

seems to consist of only short texts, it is actually a rich data source as each tweet brings with it up to 160 meta attributes, containing not only information about the tweet itself, but also about the user who wrote the tweet.

To make it easy for students to use Twitter as a data source in school, we extended the block-based programming language Snap! so that data can be retrieved from the Twitter API. To use the tool, students do not need to register with Twitter. In addition to accessing Twitter data, we also allow the use of CSV files as additional data sources, and provide map and chart functionalities for visualising data. **(HW)**

**“ TWITTER IS A RICH DATA SOURCE, AS EACH TWEET BRINGS WITH IT UP TO 160 META ATTRIBUTES**

## ACTIVITY 1: NAME AND TIMING 20 MINUTES

Give students the most simple task to get started with Snap!Twitter:

“How many meta attributes does a tweet contain in addition to its text? What can you find out about the user who wrote a tweet?”

As this task can be mastered using just two blocks, ‘tweet as table’ and ‘single tweet’, from a technical perspective it fulfils the goal of letting students explore the tool and the data source, ‘Twitter’.

The results of this activity should be summarised in plenary.

A	B
created_at	Mon Jul 29
id	
id_str	
text	@ Som
display_text_range	
source	<a href="http://twitte
user	
geo	
coordinates	
place	
contributors	

tweet single tweet as table

## OBJECTIVES

- ✓ Recognise the value of metadata
- ✓ Conduct simple data analysis
- ✓ Estimate the validity of data analysis

## ACTIVITY 2: PEOPLE USUALLY TWEET IN THEIR MOTHER TONGUE – DON'T THEY? 25 MINUTES

An interesting task for getting into data analyses is to analyse the distribution of languages over the globe. Thus, give your students the task of checking which of the most common languages (such as English, Spanish, or German) are used, and where.

```

when clicked
script variables lang colour
for each tweet
  set lang to attribute lang from tweet tweet
  if lang = en
    set colour to blue
  else
    if lang = es
      set colour to green
    else
      if lang = de
        set colour to orange
      else
        set colour to gray
  marker: position geo of tweet tweet size 5 color colour
  popup text lang
  
```

This can be done by visualising tweets on a map, distinguishing the different languages using different colours.

Obviously, the native language is not the most common language on Twitter, in most countries. Instead, English is the primary language – at least according to the 'lang' attribute.



## DIFFERENTIATION

In the lesson plan described, we've focused on students who have basic experience with block-based programming. If this isn't the case, then students will need more detailed tasks at the beginning. Experienced students should then be able to explore the tool further.

In the main activity, students can follow their own interests, either by selecting a research question that is interesting for them or by pursuing a question of their own. Since most questions at this stage cannot be fully answered in the short time available, there is always room for further improving the analyses, so even confident students have something useful to do, while those who struggle a little can at least get some basic results.

## ACTIVITY 3: LET'S ANSWER A RESEARCH QUESTION... 40 MINUTES

As the students now have some basic experiences using Snap!Twitter, accessing tweets and their attributes, and also using the map tool, they can now conduct more complex analyses. Hence, we propose to give them several questions or statements, of which each pair of students should select and investigate one. For example:

- People in Japan have fewer 'real' social contacts but are better connected in virtual life than people in the United States.

- While people in Europe typically like the colour blue, in the Asian countries people prefer green.
- Do people use more Android or Apple devices?
- In some countries, relatively few people use Twitter, while in others the usage is much higher. Which are the countries in which Twitter is most popular?

Of course, students can also follow their own ideas here!

## RELEVANT LINKS

The Snap!Twitter tool is available at: [snaptwitter.dataliteracy.education](http://snaptwitter.dataliteracy.education).

As this is released as open source software, you may set up your own server or use our server. However, for legal and technical reasons, we have to restrict access to the tool. Therefore, to get access to the Twitter area, please contact us via e-mail (see the website above).

## ACTIVITY 4: WRAPPING UP FLEXIBLE

As a last step of this lesson plan, students should think critically about the analyses they have conducted. For this purpose, all student pairs (grouped by questions, if multiple pairs worked on the same one) should briefly present their analysis, with its goals and

the way it works, to the others. In order to come into discussion about the validity of the analyses, the other students are requested to call into question whether the analysis seems reasonable, comprehensible, and if there are any weaknesses.

## DR. ANDREAS GRILLENBERGER

Andreas Grillenberger is a research fellow at Freie Universität Berlin in Germany. Since 2013, he has worked on data, data management, and data literacy topics, and investigates these from a computing education perspective, particularly with focus on secondary education.