

Health Data Privacy

Jordan Swanson

Duration: 80 minutes

Introduction

This is a free resource for teachers and students and is part of the [Callysto](#) project, a federally-funded initiative to foster computational thinking and data literacy in Canadian Grade 5-12 classrooms.

In our modern society with internet-connected devices everywhere, data is being constantly collected on us that might be more revealing than we know. Most of the time, the data is being used to target us with ads or services that we're likely to enjoy, but sometimes the data is used for less upstanding purposes.

In this lesson plan, several real-world examples are shown that go through common uses of personal data, from the good to the bad. There's an additional activity component, styled similar to a virtual 'locked room' that will allow students to discover the clues to the next stages by utilizing basic data science skills (also introduced in the notebook).

To log in to the Callysto Hub you and your students will need a Google or Microsoft account. This can be a school division-provided account or a personal account. Callysto does not collect any personal information about accounts.

Grade Level and Audience

Grade 8-12

Necessary Background Knowledge

Knowledge of Python/Pandas would be beneficial, but not required.

Learning Outcomes

- A. Show real-world examples of how data (especially health data) is used

- a. Intentionally used for public benefit
- b. Unintentionally used to reveal military secrets
- c. Intentionally used to to discriminate for health insurance purposes
- B. Give students hands-on data science experience
- C. Show how seemingly innocuous data can be revealing when combined

Required Materials

1. A charged computer.
2. Access to the internet.
3. An installed internet browser, preferably Google Chrome.
4. A Google or an Outlook email account.
5. One device per small group (2-4) of students
6. [Health Data Privacy notebook](#)

Preparation

- [Open the notebook located here](#)
- Ensure students are able to load both the notebook and the associated Google Form (also located here: <https://forms.gle/684QTAZGN3SjzmRFA>)
- Activities 1-3 could optionally be presented to the class before breaking into smaller groups
- Have students form small groups (2-4)

In-Class Activities

Activity 1: Introduction - (5 minutes)

- Read introduction paragraph
- Have students brainstorm devices that collect data, type of data collected
- Follow link in paragraph for more examples

Activity 2: Real-world Examples (20 minutes)

- Go through each of the three given examples
- Encourage students to suggest similar examples within each sub-heading
- Follow links for more information as required

Activity 3: Activity Introduction (5 minutes)

- Introduce activity from 'Data Privacy Locked Room' heading
- Carefully go over bullet points and image

Activity 4: Pandas Basics (10 minutes)

- Read through instructions with each cell
- Allow students to play around with cell contents
- Ensure at least one member of each group is comfortable with Pandas before proceeding

Activity 5: Open Google Form (1 minute)

- Follow link to open Google Form that is required to enter answers
- Instruct students to use the notebook to find clues, and enter clues in Google Form

Activity 6: Dataset #1 - Pet Group (5 minutes)

- Run the cell contents as-is to allow students to see what columns exist in the data
- Allow them to use the area code table lookup if they need a hint
 - Copy code chunk into cell
- **Solution:** Students should hopefully recognize the dog breed as a Golden Retriever OR that the only Calgary area code is 403
 - `petDf[petDf['Breed']=='Golden Retriever']` or
 - `petDf[petDf['Owner Phone Number (area code)']==403]`
 - **Answer to Google Form = 403-352-0013**

Activity 7: Dataset #2 - Internet Search Leak (10 minutes)

- Copy the code chunk from the Google Form and run the cell contents as-is to allow students to see what columns exist in the data
- **Solution:** Students should use the phone number discovered in the last section as a filtering criteria
 - `searchDf[searchDf['Phone Number']=='352-0013']`
 - **Answer to Google Form = NHANES or NHANES Study**

Activity 8: Dataset #3 - NHANES Data (15 minutes)

- Copy the code chunk from the Google Form and run the cell contents as-is to allow students to see what columns exist in the data
- **Solution:** This one requires multiple criteria to solve:
 - Jason is under 30 years old (revealed in introduction)
 - Jason is male (heavily implied in introduction)
 - Jason is married (revealed in introduction)
 - Jason has a Systolic BP greater than 140 (revealed in Google Form)
 - `nhanesDf[(nhanesDf['Age']<30) &`
 - `(nhanesDf['Sex']=='Male') &`
 - `(nhanesDf['Marital Status']!='Never married') &`
 - `(nhanesDf['Systolic BP']>140)]`
- **Answer to Google Form = 5**

Activity 9: Dataset #4 - Internet Search (5 minutes)

- Have students return to Dataset #2 (Internet search leak)
- **Solution:** Student should use *the entire phrase* as an answer to the Google Form:
 - **Answer to Google Form = spin studios near me**

Activity 10: Conclusion (5 minutes)

- Read conclusion paragraph
- Encourage students to brainstorm ways in which their data can be used negatively towards them

Reflections

Some suggested questions:

- *What went well? Why?*
- *What was tricky? How did you overcome it?*
- *What would you do differently? What would you do the same? Why?*
- *What connections can you make between what you learned from this lesson and your other coursework, lessons, or experiences?*

Next Steps

For more information, you can check out our [YouTube videos](#), [online courses](#), or [callysto.ca](#) for [learning modules](#), [tutorials](#), [lesson plans](#), [exercises](#) and events.

Contact

If you encounter any issues or have any suggestions, please get in touch with us at contact@callysto.ca or twitter.com/callysto_canada.