How often do you work with computational data analyses (eg. data science, machine learning, statistical analysis)?

- ○ Daily
- ○ Weekly
- ○ Rarely
- ○ Never

What is your occupation?

- ○ Data Scientist
- ○ Research Scientist
- ○ Student (Undergraduate)
- ○ Student (Graduate)
- ○ Post Doctoral Researcher
- ○ Professor
- ○ Other (Please specify) _____

Which field do you work in?

- ☐ Data Science
- ☐ Statistics
- ☐ Machine Learning
- ☐ Forestry
- ☐ Environmental Science
- ☐ Bioinformatics
- ☐ Other _____

Please rate your experience with the following programming languages.

	Unfamiliar	Beginner	Intermediate	Expert
Python	○	○	○	○
R	○	○	○	○

Do you have a background in computer science (ie. degree in Computer Science or related field)?

- ○ Yes
- ○ No
- ○ Other (Please explain)

Thank you for completing the Pre-Activity Questionnaire.

Press Next to move onto the Introduction and Overview Section.

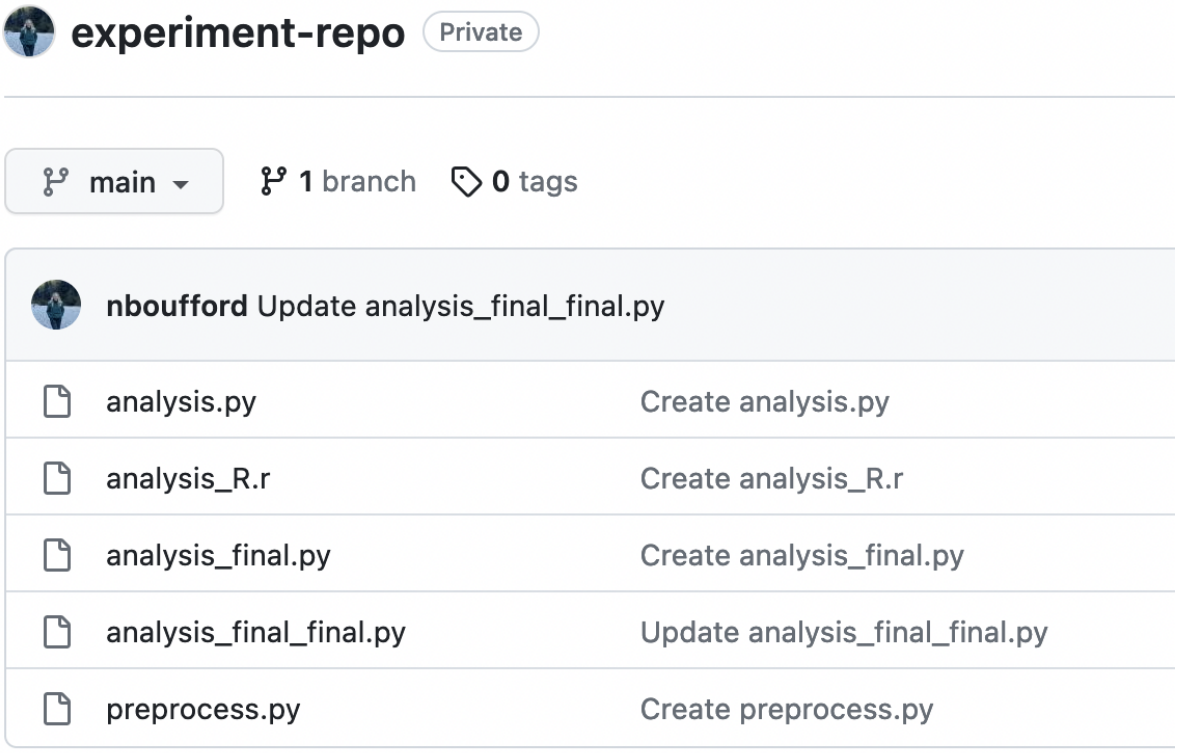
Introduction and Overview

Introduction and Overview

To help understand the purpose of this study, consider this example scenario:

Imagine you are a data science researcher who is studying the spread of wildfires using public datasets. Your supervisor has given you a promising data analysis from another researcher who has moved to another field and no longer works in your group. The data analysis uses a new model for predicting the spread of wildfires across mountainous terrain.

The researcher was kind enough to leave you an up to date repository with the analysis scripts and point you to the public data source where they got their input data. But, you look at the repository and see that there are several versions of the data analysis. Additionally, there is a pre-processing script that appears to perform data cleaning. You assume this is run before the analysis script to prep the data, but you cannot confirm. Your supervisor asks you to reproduce the previous student's results and explain the findings to the team so that you can work on next steps for this exciting research!



You first go to the public data repository and see that there are datasets spanning several years. You pick the most recent data version, although the researcher had been working on this project for a year or two already. You make a mental note to check the second most recent dataset if this one doesn't work.

Year	Dataset
2022	wildfire_data2022.csv
2021	wildfire_data2021.csv
2020	wildfire_data2020.csv
2019	wildfire_data2019.csv

You run the analysis using wildfire_data2022.csv as input, but find that you are missing some libraries that are required to run the analysis script. You cross your fingers and hope that the library versions you install on your computer are the same as the other researcher. Once you install the libraries, you run the preprocessing script, followed by the analysis script. After a few minutes, the analysis script terminates. You check the output, but none of the numbers match the results that the researcher previously reported.

```
Accuracy of the Wildfire Prediction Model: 50.00%
Expected Accuracy: 95.00%
```

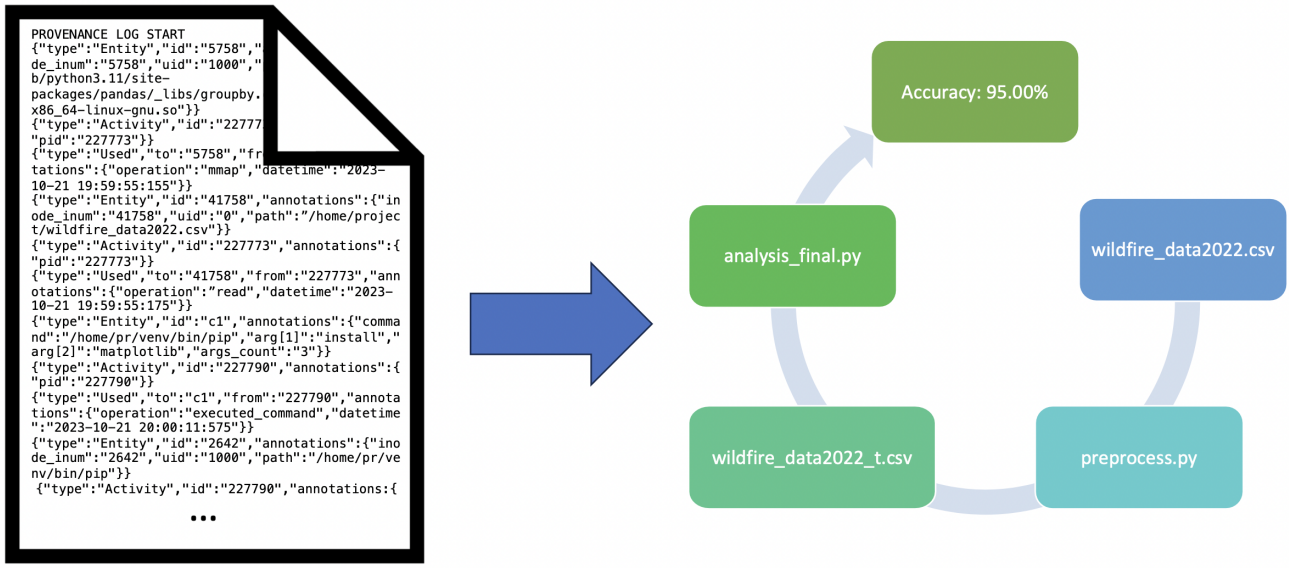
You're not sure where you went wrong, was it the correct dataset that you used? Should you have used the data preprocessing script before running the analysis? There are a few other analysis scripts in the repository that you

could try. It might be time for a coffee, because this could be a while..

But wait! You realize that, in an attempt to improve reproducibility and explainability of their research, the previous researcher had collected provenance during their experiment runs!

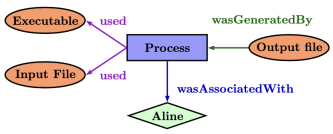
Data provenance (or just provenance) is a record of data - what has happened to the data, when it was created and modified. We can use provenance collection tools to capture provenance data during experiment execution. What we get from provenance is a record of which files, datasets, and programs were used and when. You can observe program inputs, outputs and data file versions. The provenance should tell you exactly which code and input data lead to which results as well as the computational environment!

The raw provenance log is pretty verbose and not so human readable, so you open the log in a provenance viewing application. The application summarizes the provenance and shows you the summarized provenance representation. This tool helps you to understand and reproduce the experiment easily.

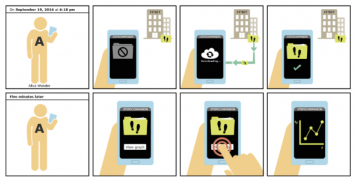


Provenance data can be shown in various ways, including logs, graphs and textual summaries.

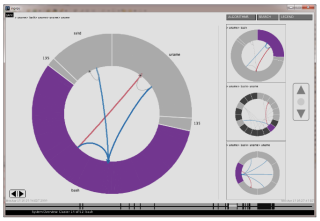
Graph Provenance Visualization [1]



Comic Provenance Visualization [2]



Circular Provenance Visualization [3]



In the following activities, we will give you a provenance data representation from a computational experiment and you will use it to answer some questions about the experiment.

We will start with a practice question to get the hang of it.

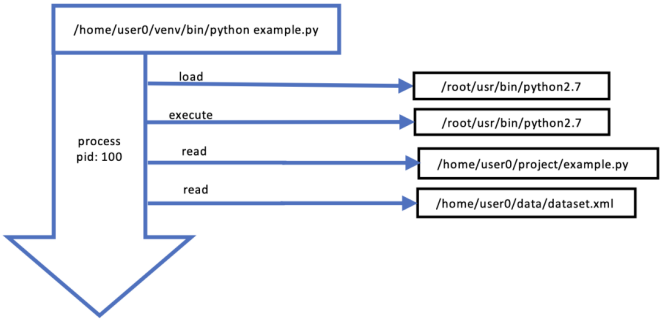
1. Pasquier, T., Han, X., Moyer, T., Bates, A., Eyers, D., Hermant, O., Bacon, J., and Seltzer, M. Runtime Analysis of Whole-System Provenance. Conference on Computer and Communications Security (CCS'18) (2018), ACM.

2. Schreiber, A., Struminski, R. Tracing personal data using comics. In: Antona, M., Stephanidis, C. (eds.) International Conference on Universal Access in Human-Computer Interaction (2017), LNCS.

3. Borkin, M., Yeh, C., Boyd, M., Macko, P., Gajos, K., Seltzer, M., and Pfister, H. Evaluation of Filesystem Provenance Visualization Tools. Transactions on Visualization and Computer Graphics (2013), IEEE.

Practice Question

The following two images are examples of a provenance summarizations of a simple computational task. The first is a graph summarization (left) and the second is a text summarization (right). Try to answer the questions below using the two provenance summarizations.



Summary:
The user executed the script "example.py", located in the /home/user0/project/ directory. They executed the script using Python version 2.7. The script accessed a dataset located at /home/user0/data/dataset.xml.

What is the name of the script being executed?

- example.py
- python3
- dataset.csv

Where is the script located?

- /home/user0/project/
- /home/user1/project/
- /home/user3/project/

Which version of Python is used when running the script?

- Python 2.7
- Python 3.7
- Python 3.8

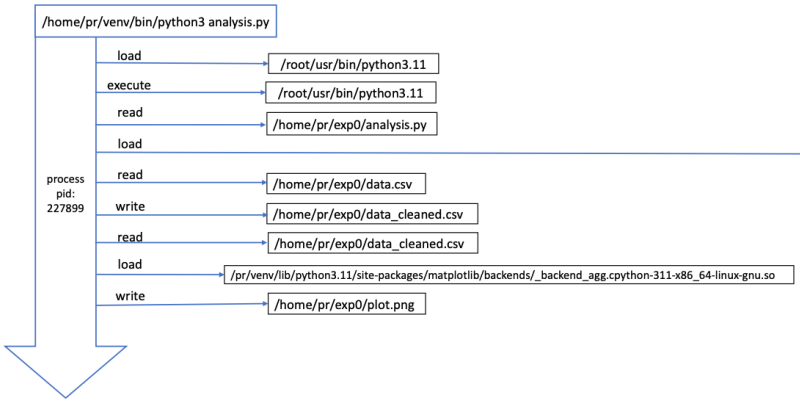
Great, you answered all of the questions correctly! Press Next to move onto the real tasks.

The following tasks will be similar to the example tasks, but we will only show either the graph summarization or the text summarization (not both!). Answer the questions to the best of your ability. After each round, you will be asked to answer several questions about the difficulty of the task before proceeded to the next task. There are 4 tasks in total.

This part of the study will be screen and audio recorded. Please let the facilitator know you are ready to move on to the study tasks so they can set-up the recording.

Task 0 (Python Script)

Please answer the following questions using the graph provenance summary.



/home/pr/venv/lib/python3.11/site-packages/numpy/core/_multiarray_umath.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy.libs/libopenblas64_p-r0-0cf96a72.3.23.dev.so
/home/pr/venv/lib/python3.11/site-packages/numpy.libs/libgfortran-040039e1.so.5.0.0
/home/pr/venv/lib/python3.11/site-packages/numpy.libs/libquadmath-96973f99.so.0.0.0
/home/pr/venv/lib/python3.11/site-packages/numpy/core/_multiarray_tests.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/linalg/_umath_linalg.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/fft/_pocketfft_internal.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/random/mtrand.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/random/bit_generator.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/random/_common.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/random/_bounded_integers.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/random/_mt19937.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/random/_philox.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/random/_pcg64.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/random/_sfc64.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/numpy/random/_generator.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/pandas_parser.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/pandas_datetime.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/interval.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/hashtable.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/missing.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/dtypes.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/np_datetime.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/conversion.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/base.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/offsets.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/timestamps.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/nattype.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/timedeltas.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/timezones.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/fields.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/ccalendar.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/tzconversion.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/properties.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/parsing.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/strptime.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/period.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslibs/vectorized.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/ops_dispatch.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/algos.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/lib.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/ops.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/arrays.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/tslib.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/sparse.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/indexing.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/index.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/internals.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/join.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/writers.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/window/aggregations.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/window/indexers.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/reshape.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/groupby.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/json.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/parsers.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/pandas/_libs/testing.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/matplotlib/_c_internal_utils.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/PIL/_imaging.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/Pillow.libs/libtiff-91af027d.so.6.0.2
/home/pr/venv/lib/python3.11/site-packages/Pillow.libs/libjpeg-32b76cef.so.62.4.0
/home/pr/venv/lib/python3.11/site-packages/Pillow.libs/libopenjp2-20e347f0.so.2.5.0
/home/pr/venv/lib/python3.11/site-packages/Pillow.libs/libxcb-f0538cc0.so.1.1.0
/home/pr/venv/lib/python3.11/site-packages/Pillow.libs/liblzma-1e44b93d.so.5.4.4
/home/pr/venv/lib/python3.11/site-packages/Pillow.libs/libXau-154567c4.so.6.0.0
/home/pr/venv/lib/python3.11/site-packages/matplotlib/_path.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/matplotlib/ft2font.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/kiwisolver/_cext.cpython-311-x86_64-linux-gnu.so
/home/pr/venv/lib/python3.11/site-packages/matplotlib/_image.cpython-311-x86_64-linux-gnu.so

Please answer the following questions using the text provenance summary.

The data scientist performed the following tasks:

1. Executed the Python script "analysis.py" using Python 3 from the virtual environment located at "/home/pr/venv/bin/python3".
2. The script was read from the file "/home/pr/exp0/analysis.py".
3. The script used several libraries from the Python 3.11 site-packages in the virtual environment, including numpy, pandas, matplotlib, and PIL (Pillow).
4. The script read data from the file "/home/pr/exp0/data.csv".
5. The script wrote cleaned data to the file "/home/pr/exp0/data_cleaned.csv" and then read from this cleaned data file.
6. The script generated a plot and saved it as "/home/pr/exp0/plot.png".

To reproduce this task, the following files are needed: "analysis.py", "data.csv", and the Python 3.11 site-packages in the virtual environment. The output files generated are "data_cleaned.csv" and "plot.png".

What is the name of the dataset the student is using?

Which directory is the dataset saved in?

- /home/pr
- /home/pr/data
- /home/pr/exp0
- /home/pr/exp0/data

What is the name of the file containing the experiment code?

How many output files are produced? (Include intermediate outputs)

- 1
- 2
- 3
- 4

Which programming languages are used to conduct the analysis in this experiment?

- Python3.11
- R
- Python2.7
- Julia
- Python3.10
- Java

Which directory is the experiment code located in?

- /home/pr
- /home/pr/exp0
- /home/pr /exp0/analysis
- /home/pr/analysis

Task 0 TLX

How mentally demanding was this task? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How hurried or rushed were you during this task? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How successful would you rate yourself in accomplishing this task? (1-Perfect, 5-Failure)

1 - Perfect	2	3	4	5 - Failure
○	○	○	○	○

How hard did you have to work to accomplish your level of performance? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How insecure, discouraged, irritated, stressed, and annoyed were you? (1-Very Low, 5-Very High)

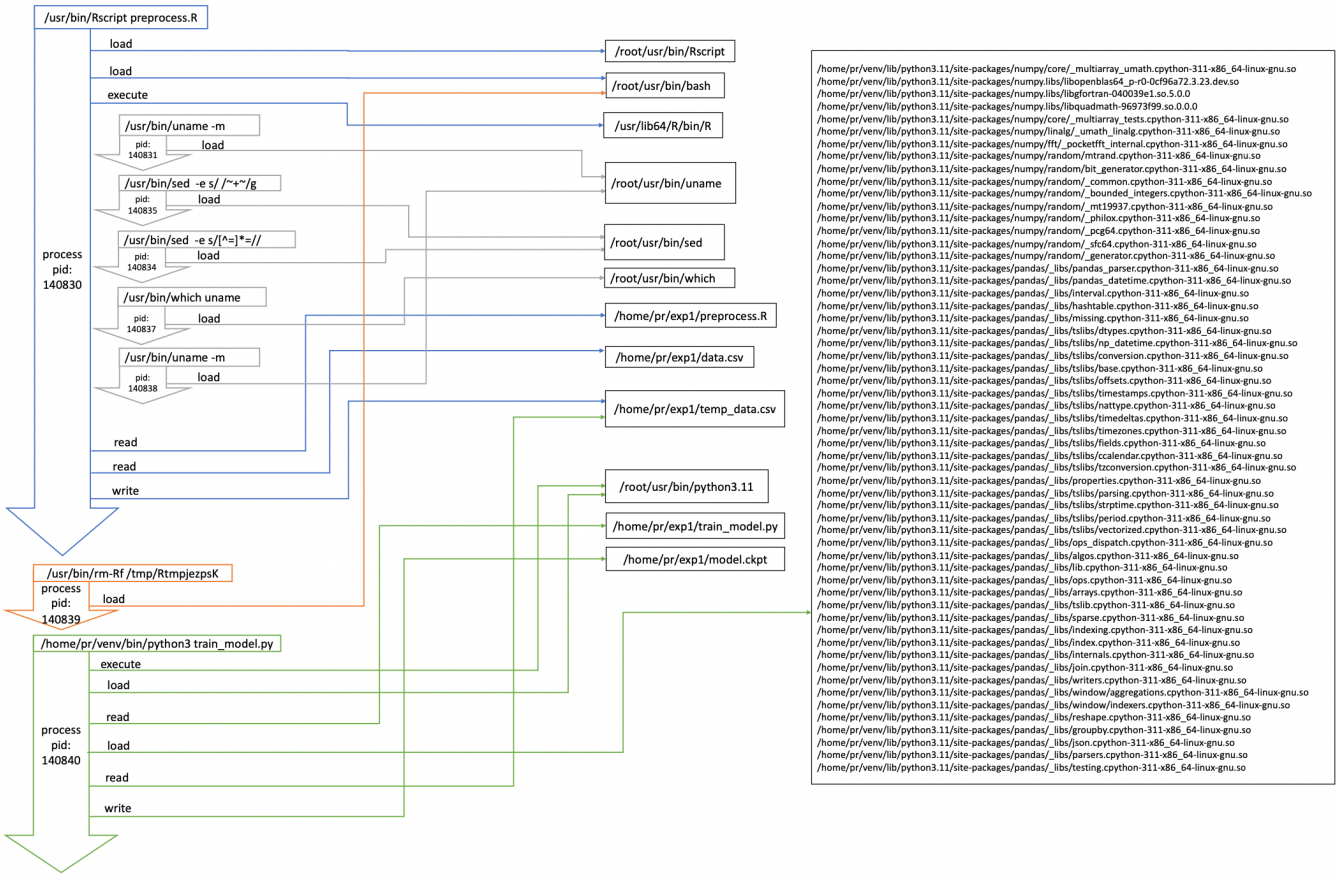
1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How useful was the provenance summary in answering the questions? (1-Very Useful, 5-Not Useful)

1 - Very Useful	2	3	4	5 - Not Useful
○	○	○	○	○

Task 1 (R and Python)

Please answer the following questions using the graph provenance summary.



Please answer the following questions using the text provenance summary.

The data scientist performed the following tasks:
1. Executed the R script "preprocess.R" using the command "/usr/bin/Rscript" and "/usr/lib64/R/bin/R". The script was located in the root directory.
2. The script "preprocess.R" was read from the location "/home/pr/exp1/preprocess.R".
3. The data for preprocessing was read from the file "data.csv" located at "/home/pr/exp1/data.csv".
4. The preprocessed data was written to the file "temp_data.csv" located at "/home/pr/exp1/temp_data.csv".
5. The temporary files created during the preprocessing were removed using the command "/usr/bin/rm" with arguments "-Rf /tmp/RtmpjezpsK".
6. The Python script "train_model.py" was executed using Python 3 from the virtual environment located at "/home/pr/venv/bin/python3". The script was located at "/home/pr/exp1/train_model.py".
7. The script "train_model.py" read the preprocessed data from the file "temp_data.csv" located at "/home/pr/exp1/temp_data.csv".
8. The script "train_model.py" used several libraries from the virtual environment, including numpy and pandas, which were located at "/pr/venv/lib/python3.11/site-packages/".
9. The trained model was saved to the file "model.ckpt" located at "/home/pr/exp1/model.ckpt".

How many times is the script "train_model.py" executed?

- ○ 1
- ○ 2
- ○ 3

How many times is the script "preprocess.R" executed?

- ○ 1
- ○ 2
- ○ 3

Which scripts **write** to the file "data.csv"?

- [ ] preprocess.R
- [ ] train_model.py
- [ ] evaluate_model.py
- [ ] None of the above

Which scripts **read** from the file "data.csv"?

- [ ] preprocess.R
- [ ] train_model.py
- [ ] evaluate_model.py
- [ ] None of the above

Which scripts **write** to the file "temp_data.csv"?

- [ ] preprocess.R
- [ ] train_model.py
- [ ] evaluate_model.py
- [ ] None of the above

Which scripts **read** from the file "temp_data.csv"?

- [ ] preprocess.R
- [ ] train_model.py
- [ ] evaluate_model.py
- [ ] None of the above

Which of the following are dependencies of train_model.py?

- [ ] matplotlib
- [ ] numpy
- [ ] pytorch
- [ ] pandas
- [ ] scikit-learn

Task 1 TLX

How mentally demanding was this task? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How hurried or rushed were you during this task? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How successful would you rate yourself in accomplishing this task? (1-Perfect, 5-Failure)

1 - Perfect	2	3	4	5 - Failure
○	○	○	○	○

How hard did you have to work to accomplish your level of performance? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How insecure, discouraged, irritated, stressed, and annoyed were you? (1-Very Low, 5-Very High)

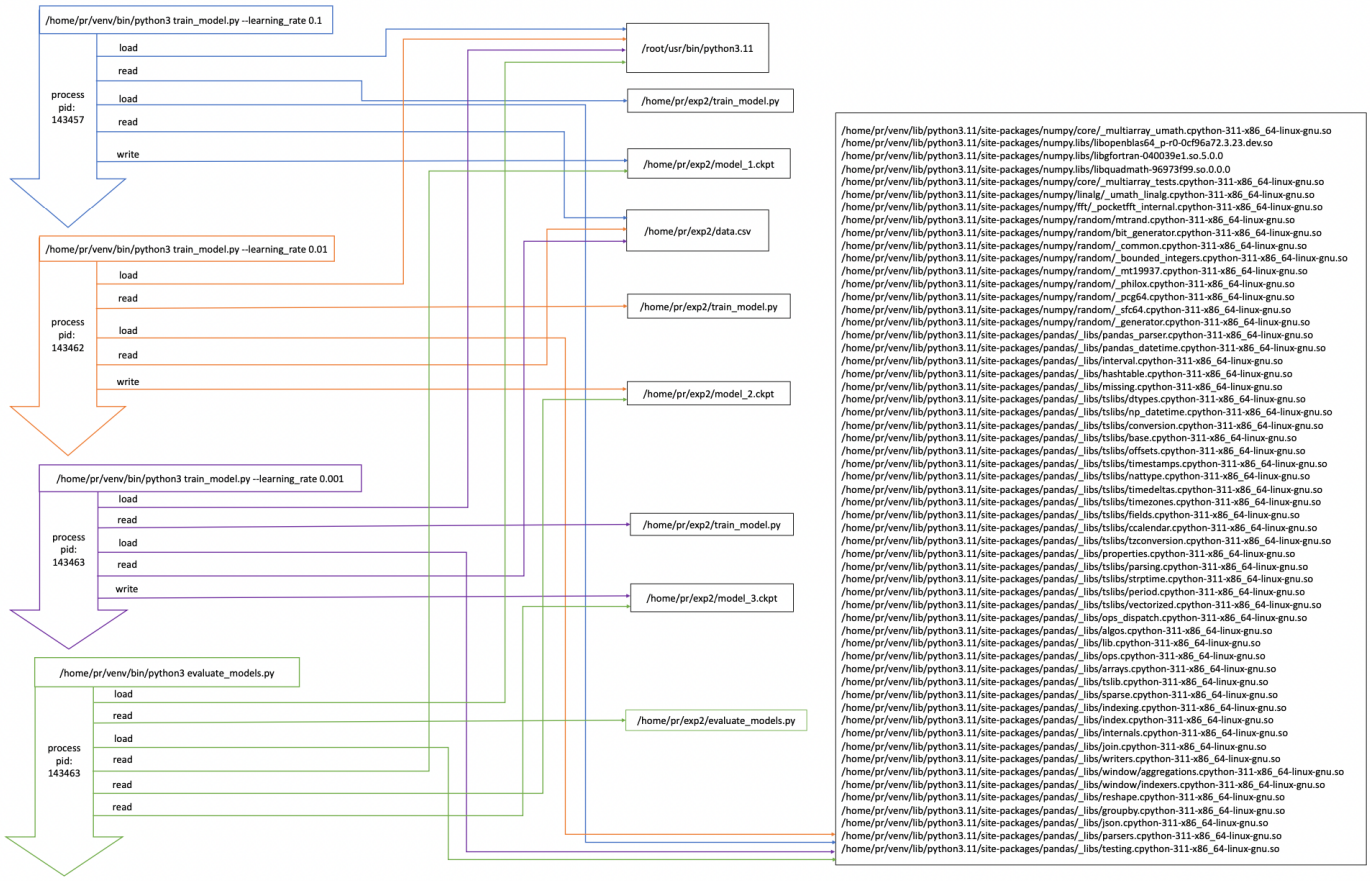
1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How useful was the provenance summary in answering the questions? (1-Very Useful, 5-Not Useful)

1 - Very Useful	2	3	4	5 - Not Useful
○	○	○	○	○

Task 2 (Command line args)

Please answer the following questions using the graph provenance summary.



Please answer the following questions using the text provenance summary.

"The data scientist performed four tasks, all involving the execution of a Python script in a Python 3.11 environment, located in the "/home/pr/venv/bin/python3" directory.

1. The first task was executed with the command "/home/pr/venv/bin/python3 train_model.py --learning_rate 0.1". The script read data from the file "/home/pr/exp2/data.csv" and wrote the trained model to the file "/home/pr/exp2/model_1.ckpt". Several Python libraries were loaded, including numpy and pandas, from the "/pr/venv/lib/python3.11/site-packages/" directory.
2. The second task was executed with the command "/home/pr/venv/bin/python3 train_model.py --learning_rate 0.01". The script read data from the same file "/home/pr/exp2/data.csv" and wrote the trained model to a different file "/home/pr/exp2/model_2.ckpt". The same Python libraries were loaded as in the first task.
3. The third task was executed with the command "/home/pr/venv/bin/python3 train_model.py --learning_rate 0.001". The script read data from the same file "/home/pr/exp2/data.csv" and wrote the trained model to a different file "/home/pr/exp2/model_3.ckpt". The same Python libraries were loaded as in the previous tasks.
4. The fourth task was executed with the command "/home/pr/venv/bin/python3 evaluate_models.py". The script read the trained models from the files "/home/pr/exp2/model_1.ckpt", "/home/pr/exp2/model_2.ckpt", and "/home/pr/exp2/model_3.ckpt". The same Python libraries were loaded as in the previous tasks.

To reproduce these tasks, the same Python environment and libraries should be used, and the "train_model.py" and "evaluate_models.py" scripts should be executed with the specified learning rates. The data should be read from the "data.csv" file, and the trained models should be saved to the "model_1.ckpt", "model_2.ckpt", and "model_3.ckpt" files, respectively. The models should then be read from these files for evaluation."

Where is the dataset located?

- /home/pr/venv/
- /home/pr/
- /home/pr/data/
- /home/pr/exp2/

Move the commands into the order they were executed.

python3 train_model.py --learning-rate 0.001

python3 evaluate_models.py

python3 train_model.py --learning-rate 0.1

python3 train_model.py --learning-rate 0.01

Do all Python executions use the same version of Python?

- Yes
- No
- Unsure

As a scientist trying to reproduce the evaluation, what are the input files you will need to run evaluate_models.py?

- data_final.csv
- model_3.ckpt
- model_2.ckpt
- new_model.ckpt
- model_4.ckpt
- data.csv
- train_model.py
- model_1.ckpt

Task 2 TLX

How mentally demanding was this task? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How hurried or rushed were you during this task? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How successful would you rate yourself in accomplishing this task? (1-Perfect, 5-Failure)

1 - Perfect	2	3	4	5 - Failure
○	○	○	○	○

How hard did you have to work to accomplish your level of performance? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How insecure, discouraged, irritated, stressed, and annoyed were you? (1-Very Low, 5-Very High)

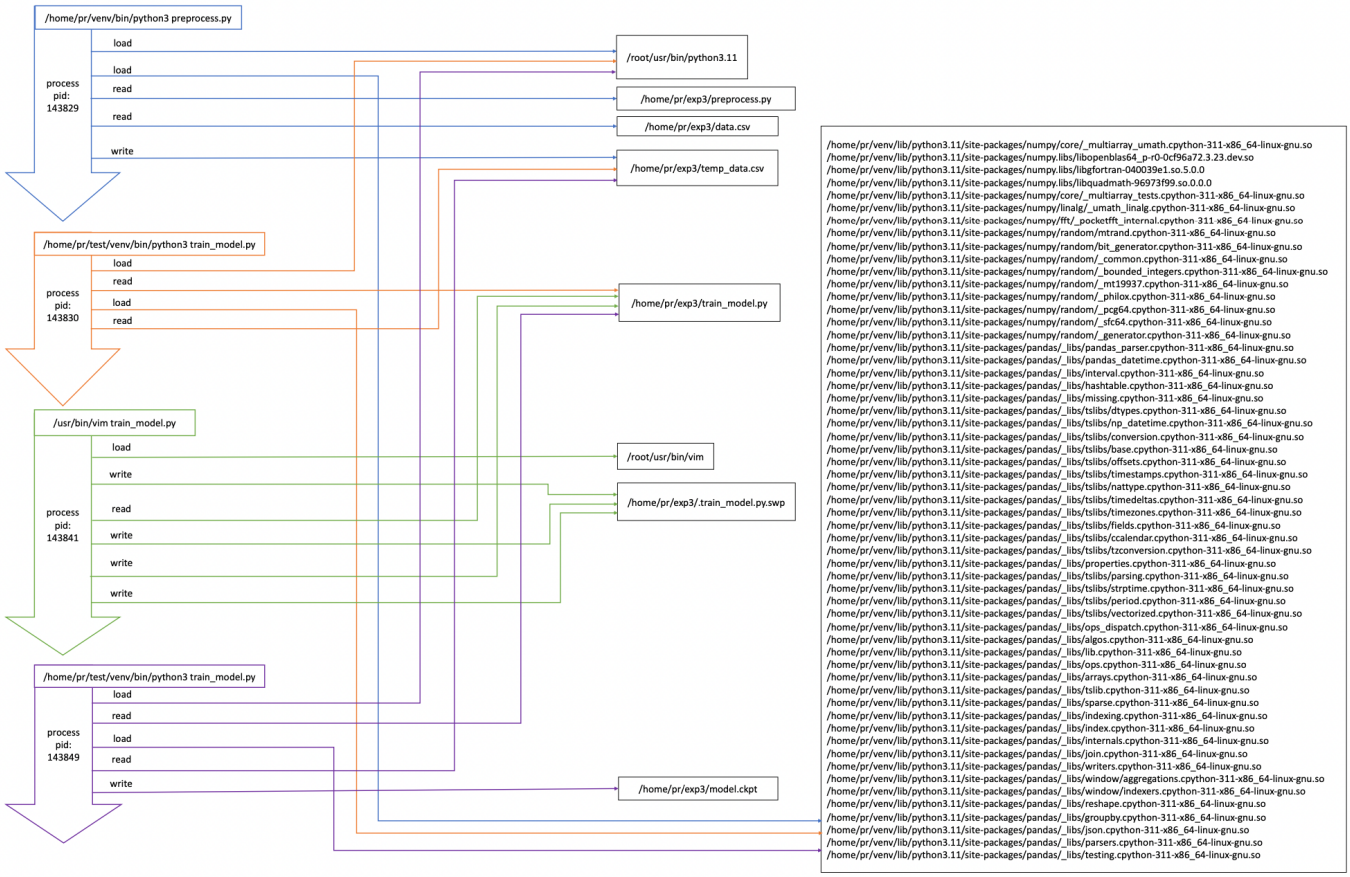
1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How useful was the provenance summary in answering the questions? (1-Very Useful, 5-Not Useful)

1 - Very Useful	2	3	4	5 - Not Useful
○	○	○	○	○

Task 3 (Preprocess, train, edit, train, evaluate)

Please answer the following questions using the graph provenance summary.



Please answer the following questions using the text provenance summary.

"The data scientist performed the following tasks:

1. Executed a Python script named "preprocess.py" using Python 3 from a virtual environment located at "/home/pr/venv/bin/python3". This script was read from "/home/pr/exp3/preprocess.py".
2. During the execution of "preprocess.py", several libraries from the numpy and pandas packages were loaded from the virtual environment's site-packages directory.
3. The script read data from a file named "data.csv" located at "/home/pr/exp3/data.csv".
4. The script wrote to a file named "temp_data.csv" located at "/home/pr/exp3/temp_data.csv".
5. Executed another Python script named "train_model.py" using Python 3 from the same virtual environment. This script was read from "/home/pr/exp3/train_model.py".
6. During the execution of "train_model.py", the same numpy and pandas libraries were loaded again from the virtual environment's site-packages directory.
7. The "train_model.py" script read data from the previously created "temp_data.csv" file located at "/home/pr/exp3/temp_data.csv". No models were saved during this process."
8. The data scientist edited the "train_model.py" script using the Vim editor. The changes were saved to the file located at "/home/pr/exp3/train_model.py".
9. The edited "train_model.py" script was then executed again using Python 3 from the same virtual environment.
10. During the execution of the edited "train_model.py" script, several libraries from the numpy and pandas packages were loaded again from the virtual environment's site-packages directory.
11. The "train_model.py" script read data from the previously created "temp_data.csv" file located at "/home/pr/exp3/temp_data.csv".
12. The script then wrote to a file named "model.ckpt" located at "/home/pr/exp3/model.ckpt". This file likely contains the trained model."

Where is the dataset located?

- /home/pr/test/venv
- /home/pr/exp/data.csv
- /home/pr/exp3/data/data.csv
- /home/pr/exp3/data.csv

How many output files were created during this experiment (including intermediate files)?

- 1
- 2
- 3
- 4
- 5

Please explain the difference between the first and second executions of the train_model.py script in no more than two sentences.

Task 3 TLX

How mentally demanding was this task? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How hurried or rushed were you during this task? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How successful would you rate yourself in accomplishing this task? (1-Perfect, 5-Failure)

1 - Perfect	2	3	4	5 - Failure
○	○	○	○	○

How hard did you have to work to accomplish your level of performance? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How insecure, discouraged, irritated, stressed, and annoyed were you? (1-Very Low, 5-Very High)

1 - Very Low	2	3	4	5 - Very High
○	○	○	○	○

How useful was the provenance summary in answering the questions? (1-Very Useful, 5-Not Useful)

1 - Very Useful	2	3	4	5 - Not Useful
○	○	○	○	○

Post-Activity Questionnaire

Overall, which provenance format was most helpful/useful when answering questions?

	Graph	Mostly Graph	Equal	Mostly Text Summary	Text Summary
	1	2	3	4	5
Use the slider to indicate preference					

Please explain your choice.

Overall, which provenance format was most enjoyable to use?

	Graph	Mostly Graph	Equal	Mostly Text Summary	Text Summary
	1	2	3	4	5
Use the slider to indicate preference					

Please explain your choice

Assuming a tool that is ready and available to use, what additional features would be useful to you in an experiment tracking tool? (eg. querying, visualizations, integration with other tools)

Assuming a tool that is ready and available to use, is there anything that would prevent you from using an experiment tracking tool, such as those described in this survey?

Is there any comments you would like to add before concluding the survey?