

Jupyter Notebooks in Education

Jeremiah W. Johnson and Karen Jin

University of New Hampshire



Abstract

Jupyter notebooks are widely used in industry for a range of tasks. This is particularly so in areas that involve significant amounts of data analysis or machine learning; indeed, while 5% of Python developers surveyed in the 2018 JetBrains Python Developer Survey report using Jupyter notebooks for their primary development tool, when restricted to those working in data science roles, Jupyter notebooks tied with the PyCharm IDE as the most popular tool for Python development [1], and in the 2019 StackOverflow developer survey, 9.5% of developers surveyed listed Jupyter notebooks as their preferred development environment [2].

Jupyter notebooks provide a format that allows the user to combine code, explanation, and analysis in a single document. The ability to mix educational or explanatory content, including, but not limited to, images, video, typeset mathematical equations, and live code makes notebooks a highly effective communication tool that enables a ‘flowing narrative’ for students to follow. This has a significant pedagogical advantage, and it is difficult to produce a similar experience in other formats. However, literature on if or how Jupyter notebooks are currently being used in education is limited, and what literature does exist is often tailored to their use in teaching specific narrow topics [3, 4]. There is little guidance in the literature on best practices for incorporating Jupyter notebooks into the curriculum.

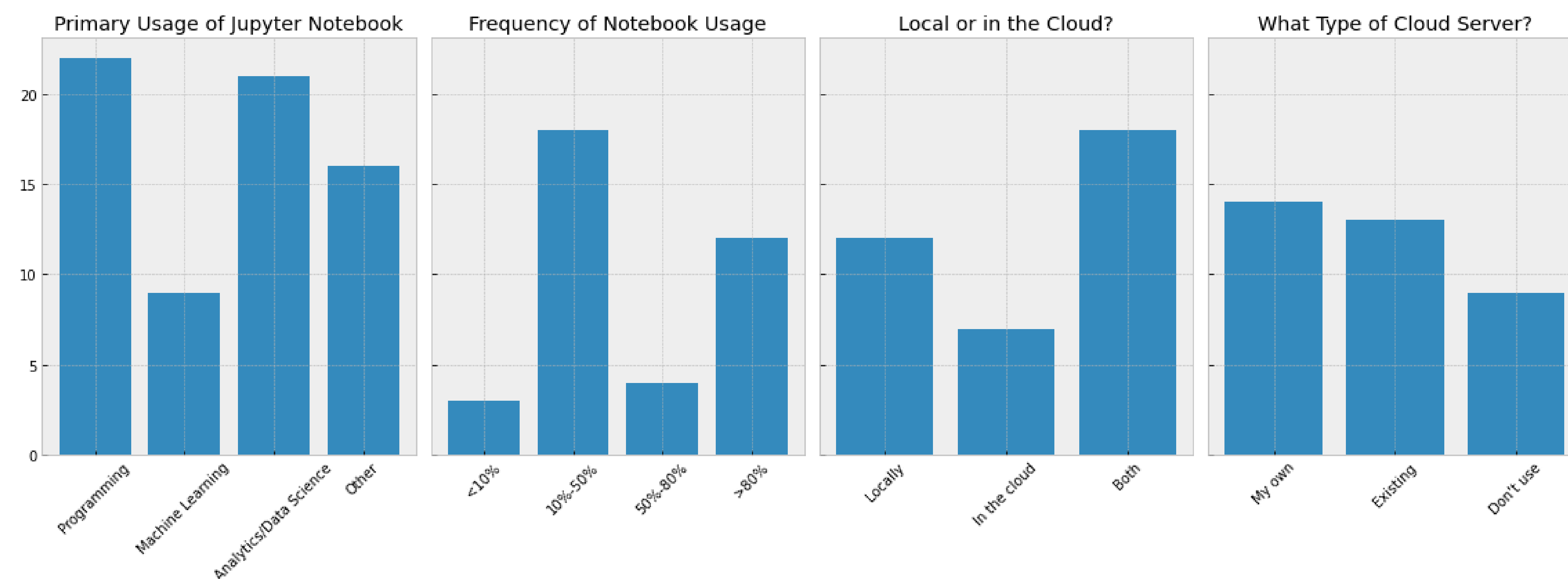
In this poster, we present the results of a survey of educators on their use of Jupyter notebooks for education. Our goal is to provide some perspective on how Jupyter notebooks are currently being used in education and to illustrate common sentiments regarding their strengths and weaknesses in the classroom, so that others considering the use of Jupyter notebooks in their courses can use them effectively.

Methodology

Over the month of October 2019 we solicited anonymous input from colleagues on our and other campuses in our college’s state system, from colleagues at various computer science education conferences in the northeast, and from participants in the online Jupyter Notebooks in Education Google group. We received 37 responses to the survey in total from educators who currently use Jupyter notebooks in one or more of their courses.

The survey consisted of seven questions, the first of which simply confirmed that the survey participant was an educator who used Jupyter notebooks in one or more of the courses that they teach. Questions 2–5 required responses within the predefined categories of the survey and serve primarily to provide background and context for the responses received to questions 6 and 7, which were open response. In addition, in question 2 (‘primary usage of the Jupyter notebook’), respondents were asked to provide uses that fell outside the categories defined by the survey question.

Survey Results



A Sample Notebook



Figure: An example of a typical Jupyter notebook showing a

Survey Questions

- What types of courses do you use Jupyter notebooks in?
- How frequently do you use Jupyter notebooks in the courses that you teach?
- Where do you run the notebooks that you use for education?
- If you run notebooks in the cloud, do you create your own cloud environment, or use an existing service such as Google Colab?
- In your view, what are the advantages of using Jupyter notebooks for teaching?
- In your view, what are the disadvantages of using Jupyter notebooks for teaching?

Discussion

The primary reason that survey respondents gave for choosing to use Jupyter notebooks in an educational setting is that notebooks provide the ability to integrate narrative content, including formatted equations, images, and plots into a single cohesive document with executable and editable code samples. However, the narrative advantage of the Jupyter notebook comes at a cost. Survey respondents noted frequently that Jupyter notebooks present a challenge to novice programmers. While local installation challenges can be mitigated by running a cloud-based notebook server, this places an additional burden on the instructor, who must then maintain the server. Another significant concern for those teaching software development is that the Jupyter notebook is a difficult environment for students to learn and apply good software engineering practices, such as modular code design or test development. The hidden state issue [5] is perhaps the most vexing issue for those using Jupyter notebooks for educational purposes (c.f. Figure 2).

Hidden State

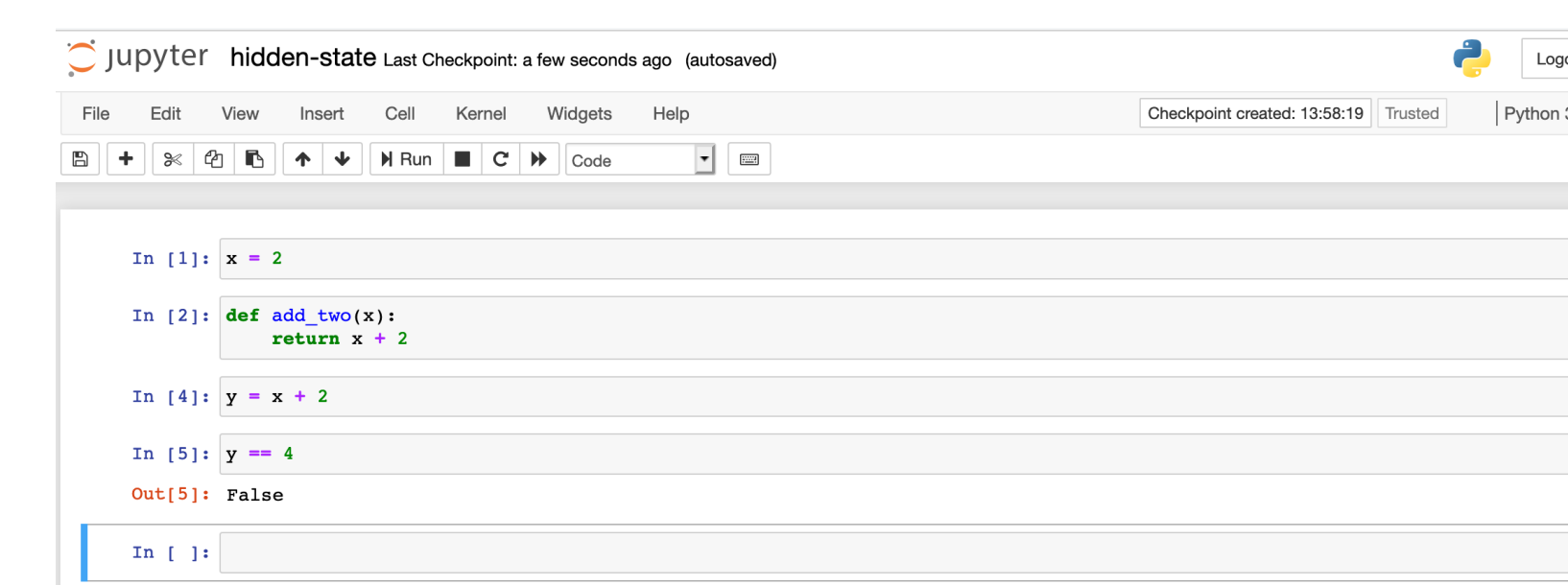


Figure: A simple example illustrating the pitfalls of hidden notebook state. In this case, a third cell was executed after the cell containing the function definition and was then deleted. Hidden notebook state was considered by survey respondents to be particularly problematic for novice users.

Conclusions & Future Work

- Jupyter notebooks are currently being used to teach a range of courses primarily, but not exclusively, in computing and mathematics.
- Jupyter notebooks are ideal for combining code, explanation, and analysis in a single document, providing a clean and coherent narrative that has a significant pedagogical advantage and is difficult to produce in other formats.
- Jupyter notebooks can be difficult for novices to install and successfully get started with.
- Jupyter notebooks maintain a hidden state that affects all notebook cells. This can cause significant challenges for novices.

The results in this survey indicate that many instructors are finding a place for the Jupyter notebook in their classrooms, despite the potential challenges that they pose for novice programmers. The results presented here (originally intended for publication at the 2020 CCSCNE conference; cancelled due to COVID-19) provided a basis for a more in-depth investigation and establishment of a set of best practices for teaching using Jupyter notebooks that was published by the first author late last year [6]. Refining and improving these guidelines is an area of ongoing research.

References

- [1] JetBrains 2018 developer survey.
<https://www.jetbrains.com/research/python-developers-survey-2018/>. Accessed 25 Nov. 2019.
- [2] StackOverflow 2019 developer survey.
<https://insights.stackoverflow.com/survey/2019>. Accessed 25 Nov. 2019.
- [3] Roland DePratti.
Using Jupyter notebooks in a big data programming course.
J. Comput. Sci. Coll., 34(6):157–159, April 2019.
- [4] Ben Glick and Jens Mache.
Using Jupyter notebooks to learn high-performance computing.
J. Comput. Sci. Coll., 34(1):180–188, October 2018.
- [5] Joel Grus.
I don’t like notebooks.
<https://docs.google.com/presentation/d/1n2RlMdmvlp25Xy5thJUhkKGvjtV-dkAIsUXP-AL4ffI/edit?usp=sharing>. Accessed 25 Nov. 2019.
- [6] Jeremiah W Johnson.
Benefits and pitfalls of jupyter notebooks in the classroom.
In Proceedings of the 21st Annual Conference on Information Technology Education, pages 32–37, 2020.