Hands-on Assignments for Practical Data Science Education to Non-Computing Majors

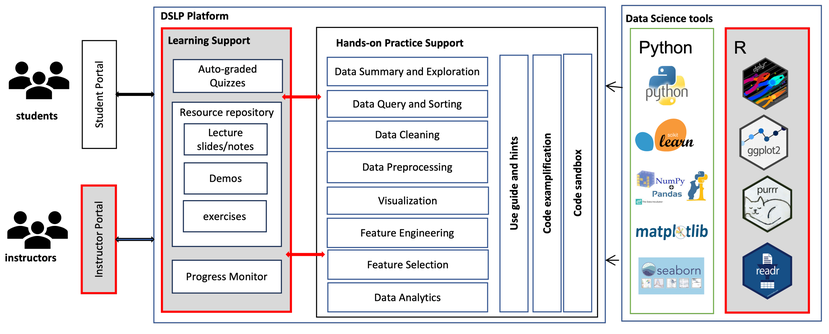
Xumin Liu, Erik Golen
Rochester Institute of Technology

Overview

- This work is part of the result of the NSF projects (DUE #2021287): Developing a Hands-on Data Science Curriculum for Non-Computing Majors.
- Hands-on practice is an essential component in data science education.
- It is challenging to offer in-depth hands-on practice to students with limited or no programming background, such as k-12 and non-computing majors.
- We designed two types of hands-on assignments in an entry-level data science course for non-computing majors at RIT.
  - Google Colab assignments: write Python Code with the support of sample code
  - DSLP assignments: perform data query, cleaning, exploration, visualization, and analysis on a web-based platform
- We conducted a survey on how effectively the two types of hands-on assignments helped students improve their interest and knowledge in data science.

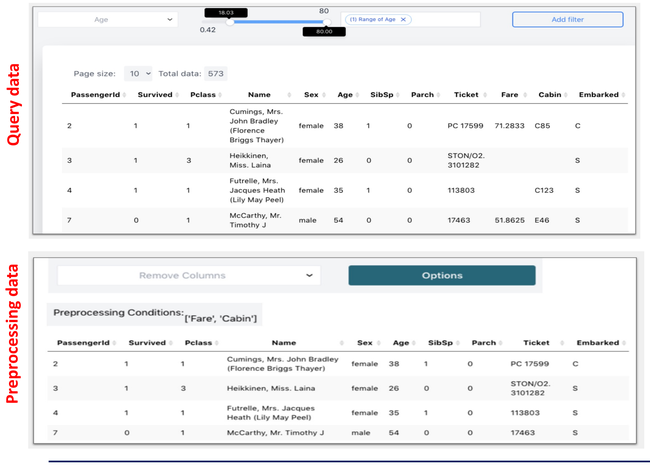
Data Science Learning Platform

- DSLP is a **Web-based p**latform that allows users to perform data science tasks **without the need for coding**

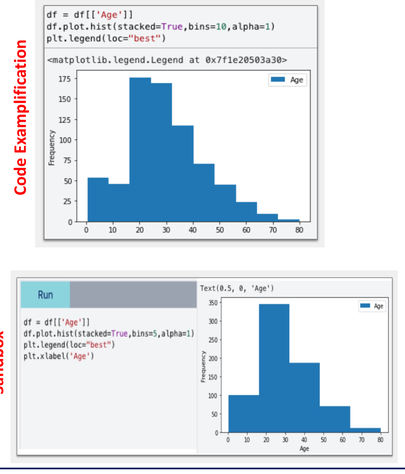


- Non-Coding interface for process/analyze data
- Cover the entire data science pipeline
- Multiple levels of coding support
- Informative instructions
- Platform usage tracking
- Extensible and flexible deployment options

GUI to work with the data

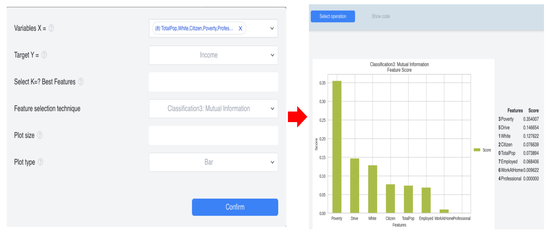


Supports in learning coding



Sample DSLP assignment:

- Questions about feature selection:



- Based upon the ranking result, should we choose Fare or Pclass to continue with in our analysis?
- Should we continue to use the Embarked feature in our analysis? Why or why not?

Sample Google Colab assignment:

- Questions about feature selection:
- Use sklearn.feature_selection.SelectKBest to list the scores of each attribute, choose a meaningful score_func using the table in slide 21.
- The following sample code is given

```
from sklearn.feature_selection import SelectKBest, mutual_info_regression #don't forget to import the score function
k=5 # top 5 features
fit=SelectKBest(mutual_info_regression, k).fit(feature, target) # here "feature" should be the one after applying the
result=pd.DataFrame({'Features': feature.columns, 'Score':fit.scores_}) # this line of code puts the content of the r
result.sort_values(by='Score', ascending=False, inplace=True) # this line of code sorts the features based on the rel
print(result) # print out the ranking result
```

Evaluation result

- **Focus**: how effectively the assignments helped students (1) understand data science principles and practices; (2) improve their self-efficacy about, and interest in, data science and computer science.

Student Background & Demographics

- Total participants: 26 students in ISCH-370: Principles of Data Science.
- Majors: EE, Econ, ME, Image Science, BioChem, Motion Picture Science, ISE, History, Biomedical, MIS, Public Policy, Political Science, Film Production Live Action, Accounting, and others.
- Year of program: 4 in 2nd year, 6 in 3rd year, 10 in 4th year, 5 in 5th year, 1 not specified.
- Demographics (choose all that apply): 21 male/ 5 female; 8 Asian/2 African American or Black/21 White/1 Native Hawaiian or other Pacific Islander / 1 as other

Survey result

Questions		A or SA	N	D or SD
Q1: Easy to Use	Colab	60%	30%	10%
	DSLP	**85%**	5%	10%
Q2: Improve Understanding	Colab	75%	20%	5%
	DSLP	**80%**	10%	10%
Q3: Increase Interests	Colab	55%	20%	15%
	DSLP	40%	50%	**10%**
Q4: Improve self-efficacy	Colab	75%	20%	5%
	DSLP	70%	25%	**5%**

Survey summary

1. Students have very diverse backgrounds and demographics. The sample size is limited.
2. Both types of assignments received positive feedback from the majority of students on effectively improved student understand, interests, and self-efficacy.
3. DSLP assignments seemed to perform significantly better than Colab ones in terms of ease of use and improving student learning.
4. Colab assignments got more votes for A or SA than DSLP ones in terms of improving interests and self-efficacy. However, it received more D or SD votes as well.

Conclusion

- Using a public software tool, such as Google Colab, might better improve student interest and self-efficacy. However, students need intensive support in doing Colab assignments (explicitly specify the function needed and provide the sample code). The variety and depth of the assignments could be limited.
- DSLP can be a better option of providing hands-on experience to students with no or limited coding background.
- DSLP provide a flexible support for student learning need. With it, students can focus on data science topics without worrying about how to code. Students can also use it to generate sample code and try their own code.

For more info

- Link to the project: https://cs.rit.edu/~xl/IUSE.html
- Link to the DSLP: http://dslp.cs.rit.edu:8000/

Project team

- **Project investigators:** Xumin Liu, Erik Golen, GCCIS
- **Research assistants**: Liyi Zhang, Sophie Lingmin Hou, Xinyu Zhang, Eduardo Lima, Vivek Gupta, Maggie Long, Shaun Pinto, Chenghui Zhu, Abhinab Acharya, Andrew Wu, Dingrong Wang, Brian Lin, Patrick Johnson, Shimanto Bhowmik.