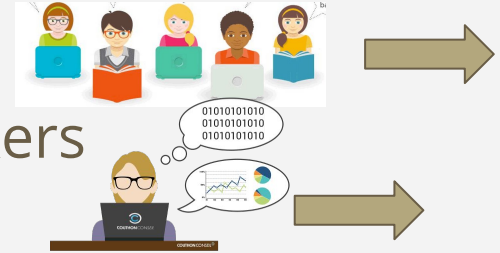
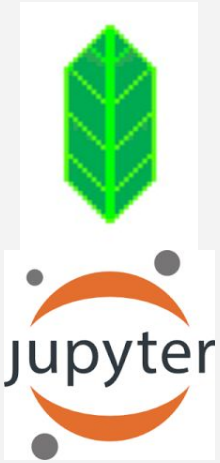
PISAnalysisTool

Nixi Wang, Hongbin Qu, Luyu Xu, Shenghao Xie

User cases

- Teachers, students, and parents
- Education researchers, policy makers and data analysts





11219	PV6SSES	Plausible Value 6 in System Subscale of Science - Earth	NUM	8.3	915	0 - 934.21
11220						
11221	PV7SSES	Plausible Value 7 in System Subscale of Science - Earth	NUM	8.3	916	34.017 - 922.825
11222						
11223	PV8SSES	Plausible Value 8 in System Subscale of Science - Earth	NUM	8.3	917	34.746 - 935.05
11224						
11225	PV9SSES	Plausible Value 9 in System Subscale of Science - Earth	NUM	8.3	918	12.619 - 928.349
11226						
11227	PV10SSES	Plausible Value 10 in System Subscale of Science - Eartl	NUM	8.3	919	16.907 - 898.335
11228						
11229	SENWT	Senate Weight (sum of 5000 per country)	NUM	12.5	920	0.00375 - 12.89587
11230						
11231	VER_DAT	Date of the database creation	CHAR	$20.	921	
11232						
11233						

Student - QQQ | Student - QQ2 | School | Cognitive | Timing | Teacher | Financial Literacy | Coll. Pb. Solving

PISAnalysisTool

Installation

`PISAnalysisTool` for its first version 0.1.0 is only compatible with Python 3.

Option 1

```
pip install PISAnalysisTool
```

Option 2

Conda env and pip requirement.txt

1. `git clone https://github.com/nixiwang/PISAnalysisTool`
2. `cd PISAnalysisTool/`
   `python setup.py install`
3. Create an environment from an exported env:

   `conda env create -f environment.yml`

   run `pip install` the following command:

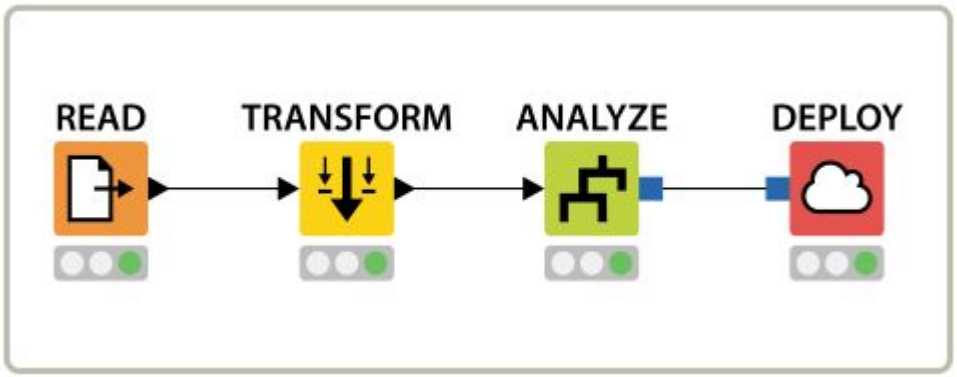
   `pip install -r requirements.txt`

Multiple data sources

PISA: 2015

- students.csv
- schools.csv
- teachers.csv

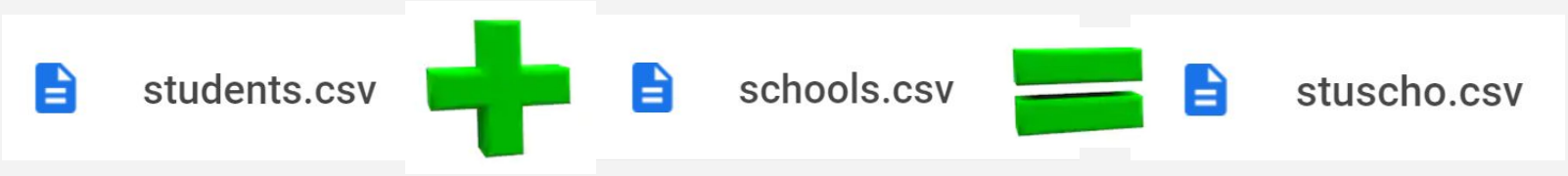
WB: import wbdata (demo)



Four components:

Data retrieving and cleaning

- merge_csv.py
- Takes a student/teacher file as a parameter
- Merges the file with the school file based on:
  - ‘CNTSCHID’: School ID
  - 'SUBNATIO': Nation ID



PISA Data processing

- get_students_info.py
- Takes merged school-student file and user-specified attributes as inputs
- Extracts user-specified attributes and generates a new csv file for further modeling and visualization

http://localhost:8888/notebooks/pisa_analysis_tool/get_student_info.ipynb

Hierarchical generalized linear model

1. **Exemplar inquiry:**

   *** Gender difference in science assessment internationally***

   - What school and country-level factors are contributing to science performance?
   - What sociocultural factors are mediating the gender effect?

2. Demo for running main function and specific function lookup:

Jupyter notebook demo

User cases: customized model functioning

Demo:

- Importing our `formula_creator` module
- <http://35.247.97.57:8888/notebooks/Notebooks/demo%20formula%20generator.ipynbk>

Visualization

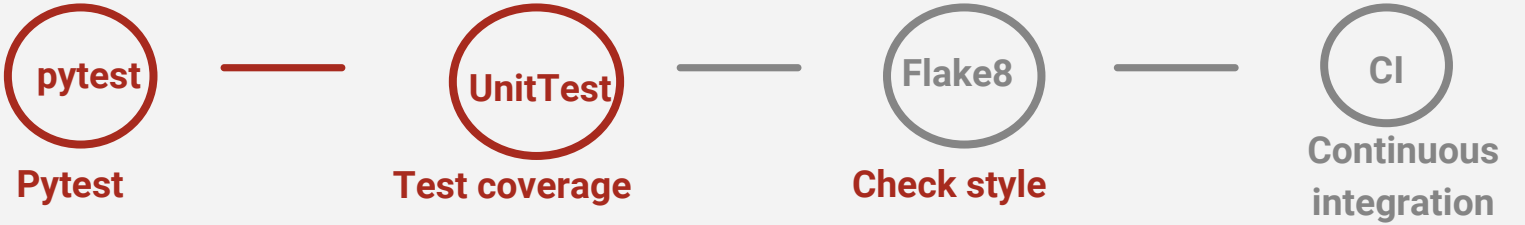
http://35.247.97.57:8888/view/Notebooks/map_data.html

Visualization

Gender test difference coefficient: $\frac{Male\ Score - Female\ Score}{Male\ Score} * 100\ \%$

Testing

All tests passed!



Pytest

Run pytest for all modules individually

Test coverage

pytest mymodule
•coverage run –m unittest discover

Check style

Check styles for PEP8 and flake 8

Continuous integration

Build everything through travis continuation, all passed!

Project structure

- ‘Pisa_analysis_tool’ package folder (unit tests included)
- ‘doc’ folder
- ‘Examples’ folder
- README.md
  - https://github.com/nixiwang/PISAnalysisTool
- LICENSE
- requirements.txt
- environment.yml
- Documentation, using autodoc generation system (Sphinx)
  - file:///Users/nixiwang/git-practice/PISA-and-World-Bank-Educational-Data-Project/doc/gen/py-modindex.html
- setup.py for PyPi installation
  - https://pypi.org/project/PISAnalysisTool/(need to be updated)

Lessons Learned

- Designing user cases and components interaction
- Cultivating good coding style
- Writing elaborate unit tests
- Using version control tool
- Familiarizing some python visualization packages eg Folium
- Valuable experience on working in a team to conduct a software development project

Future work

- Refine component and function specifications
- Improve the automation level of formula creator
- Provide more forms of visualization plots
- Add more documentation to PyPI, check PEP 8 compliance
- Develop more generic functions
    - ICCs
    - Comparing model fits

Thank you

Question & Discussion

Nixi Wang, Hongbin Qu, Luyu Xu, Shenghao Xie