Was Clinton Right? -- CS 1420 Midterm Version

Read the instructions carefully. Not following instructions will result in you not earning the score you want.

Learning Outcomes

1. Use techniques you have learned so far in the class to produce a solution.
2. Read and process data in files.
3. Create data in a form that is convenient for processing by a program.
4. Formulate an argument and draw a conclusion of your own based on data.
5. Interpret data you are given so that you can answer a question.
6. Work with spreadsheet data (CSV text file) in a Python program.

Problem

Former President Bill Clinton said in his speech to the Democratic National Convention in Charlotte in 2012:

"Since 1961, for 52 years now, the Republicans have held the White House 28 years, the Democrats 24. In those 52 years, our private economy has produced 66 million private-sector jobs. So what's the jobs score? Republicans 24 million, Democrats 42 (million)."

In the packed convention hall, it was one of the night's biggest applause lines.

Did he have his facts right? Your project is to convince the grader whether Clinton was right or wrong.

We provide you a formatted copy of the file BLS_private.csv which is a download from the from the Bureau of Labor Statistics (bls.gov). Data in the file are in thousands. Clinton probably got his data from an article on bloomberg.com that is no longer publicly available. Bloomberg's writer got their data from the Bureau of Labor Statistics (bls.gov) at the U.S. Department of Labor—the definitive, non-partisan source of jobs data. In the quote above you can see that he refers to private employment data, "seasonally adjusted". The data can be found here (http://data.bls.gov/cgi-bin/surveymost?ce) if you check the box "Total Private Employment." You then need to select the years starting in 1961 up to 2012. The result is an MS Excel file that can be saved in comma-separated values format (csv). You should verify that we are providing you the correct data.

You need data on the presidents in the specified time period so that you can check Clinton's statement. What data do you need? Part of your task is to figure out what data you need. You will have to research the presidents to gather data and store it in a file that you create called presidents.txt. Your presidents.txt file cannot have any jobs data in it-all jobs data must come from the BLS_private.csv file. Your presidents.txt file will simply be a file of words and numbers that you can create using a simple text editor or Excel.

Your task is to analyze and synthesize the input from the BLS_private.csv and your presidents.txt file to see if you can verify or discredit Clinton's statement.

There are a number of implicit challenges such as "What does Clinton mean by 'produced'?" and other ideas open for interpretation.

You may ask what output we expect from your program. That is purposefully unspecified--you decide what is the most effective way to make your point. You need enough output to show that YOU have verified or discredited Clinton's assertion. Simply saying "verified" (or not) is not sufficient: justify your assertion. Simply quoting other articles will earn you 0 points, regardless of effort. Your code, your comments, and your output need to convince the grader that *YOU* solved the problem with understandable code that answered the question being asked.

You will write and/or modify several files as part of this project. Put all of these files in a directory called `clinton`:

- `BLS_private.csv` contains the provided data. You may modify this file. If you ever need the original, you can find it on Canvas.
- `presidents.txt` should contain information that your analysis can use about presidents. This file must not contain jobs data!
- `fact_check.py` will contain your implementation code. Conduct analyses of the data. Print results to `conclusions.md` when run.
- `conclusions.md` should contain your answers (in prose!) to the questions. You may enter plain text or (optionally) use markdown formatting (https://daringfireball.net/projects/markdown/syntax). You may write the contents of this file manually or programmatically or a combination of both. `check50` will expect you to submit this file.

The provided BLS csv data file is cleaned up, and should be straightforward to use with any of the various python modules you might consider using. You choose which data structures you want to use in analyzing your data-that is not specified for you. However, we do not expect you to use classes. Note that two presidents did not complete their terms and that may change the numbers a bit. A correction might affect the results, but not significantly.

Bonus Questions

As the saying goes, "There are three kinds of lies: lies, damned lies, and statistics." (attribution is complicated (https://en.wikipedia.org/wiki/Lies,_damned_lies,_and_statistics)). Simply having correct data doesn't mean that you are telling the correct story, and that could apply to Clinton's quote. In what ways is this the wrong data for the context in which it was being used? How could the numbers be adjusted to better reflect the influence that a president has on this data?

Answer these questions for bonus points.

How to Test

- Run your program on your possibly modified labor data input file and using your `presidents.txt` file. Verify that the output is what you expect.

- Any output you have, whether text or pictures, should be to files. Make sure that your program terminates correctly.

Correctness

In your terminal, execute `ruff check` to check for syntax errors, violations and code that leads to potential bugs.

```
ruff check fact_check.py
```

Most aspects of your code must be checked manually.

Style

Execute the below to evaluate the style of your code using `ruff`.

```
ruff format fact_check.py
```

How to Submit

From your assignment repository page in GitHub Classroom, click "Commit" button and enter a meaningful commit message.

Frequently Asked Questions

Q: What does it mean that this project is the midterm?
A: It means that it counts in your course grade as the midterm. In CS 1400 it is the final exam.

Q: Since this is an exam, can we help each other like we have done on the other projects? Can we get help from our instructor or the Lab TAs? How about Copilot?
A: You can ask clarifying questions, but do not expect as much help as on other projects. In the words of Yoda "Demonstrate all that you have learned.". Use Copilot as you have been taught to use it.

Q: Who creates the `presidents.txt` file? What format is it?
A: You create it. It's a simple text file. You put in it whatever data you think will help you answer the question. Just no job numbers data allowed.

Q: Does the `presidents.txt` file have to be a CSV file?
A: No. It can be, if you want it to.

Q: I can answer this question using Excel. Why do we need a Python program to do this?
A: The simple answer is automation. Yes you could figure out a way to do this in Excel without much programming once. But what if you wanted to answer this kind of question over and over? Doing it manually is tedious and error prone.

Q: Why are you asking us to verify the input data that you give us? Is this a trick?

A: Any time someone makes a claim based on data they give you, you should attempt to verify that the data is correct and hasn't been tampered with.

Q: I think I'm doing everything right but my numbers don't add up. I don't see how Clinton could possibly have been right, though I'm not saying he was or wasn't. Any suggestions?

A: Review the description. Question your assumptions about the data and the question and the problem.

Q: Personally, I don't like President Clinton, so he must have been wrong. Why did you ask such a controversial question?

A: Controversy is part of what makes the problem interesting. Your feelings about him as a person or a president have nothing to do with data and facts. Answer the question based on the data.

Q: Personally, I think President Clinton was a good president, and he must have been right. Why did you ask such a controversial question?

A: Controversy is part of what makes the problem interesting. Your feelings about him as a person or a president have nothing to do with data and facts. Answer the question based on the data.

Q: Is there a right answer?

A: Yes there is, but this project is about much more than just the answer.

Grading

Criterion	100 (Mastery)	85 (Proficient)	70 (Basic)	50 (Needs Improvement)
Labor data used as input is correct and complete	All labor data is accurate and complete, with no missing or incorrect entries.	Most labor data is accurate and complete, with minor errors or omissions.	Some labor data is accurate, with noticeable errors or omissions.	Significant errors or omissions in labor data.
Give results for democratic presidents	Results for all Democratic presidents are correctly calculated and presented.	Results for most Democratic presidents are correctly calculated, with minor errors.	Results for some Democratic presidents are correctly calculated, with noticeable errors.	Significant errors in results for Democratic presidents.
Give results for republican presidents	Results for all Republican presidents are correctly calculated and presented.	Results for most Republican presidents are correctly calculated, with minor errors.	Results for some Republican presidents are correctly calculated, with noticeable errors.	Significant errors in results for Republican presidents.

Criterion	100 (Mastery)	85 (Proficient)	70 (Basic)	50 (Needs Improvement)
Give a clear conclusion that answers the question being asked	Conclusion is clear, concise, and directly answers the question.	Conclusion answers the question but lacks clarity or conciseness.	Conclusion is somewhat clear but lacks directness.	Conclusion is unclear or does not directly answer the question.
Results are correct given assumptions made	All results are correct based on the stated assumptions.	Most results are correct, with minor errors based on the stated assumptions.	Some results are correct, with noticeable errors based on the stated assumptions.	Significant errors in results based on the stated assumptions.
Conclusion drawn is reasonable given assumptions made	Conclusion is logical and reasonable based on the assumptions.	Conclusion is mostly logical, with minor issues based on the assumptions.	Conclusion is somewhat logical, with noticeable issues based on the assumptions.	Conclusion is illogical or unreasonable based on the assumptions.
The correct answer is given (yes, there is one)	The correct answer is provided with no errors.	The correct answer is provided, with minor errors.	The correct answer is provided, with noticeable errors.	The correct answer is not provided or contains significant errors.
presidents.txt file contains no labor data, only data about presidents	The file contains only data about presidents, with no labor data.	The file contains mostly data about presidents, with minor labor data entries.	The file contains some labor data entries.	The file contains significant labor data entries.
There is a module docstring with the required information, including exactly how to run this program	The module docstring is complete and includes all required information.	The module docstring is mostly complete, with minor omissions.	The module docstring is somewhat complete, with noticeable omissions.	The module docstring is incomplete or missing required information.
The program validates inputs and fails gracefully when necessary	The program validates all inputs and handles errors gracefully.	The program validates most inputs and handles errors, with minor issues.	The program validates some inputs and handles errors, with noticeable issues.	The program fails to validate inputs or handle errors gracefully.
Code has a main function with conditional execution. There is no global code	The code has a main function with conditional execution and no global code.	The code mostly follows this rule, with minor global code.	The code somewhat follows this rule, with noticeable global code.	The code has significant global code.

Criterion	100 (Mastery)	85 (Proficient)	70 (Basic)	50 (Needs Improvement)
Every function has a proper function docstring	All functions have proper docstrings.	Most functions have proper docstrings, with minor omissions.	Some functions have proper docstrings, with noticeable omissions.	Significant omissions in function docstrings.
Passes ruff checks for syntax errors, violations, and potential bugs	Code passes all ruff checks for syntax errors, violations, and potential bugs.	Code passes most ruff checks for syntax errors, violations, and potential bugs.	Code passes some ruff checks for syntax errors, violations, and potential bugs.	Code does not pass ruff checks for syntax errors, violations, and potential bugs.

Students should strive for mastery level. Lower than that indicates areas where more practice is needed or more learning is needed. Code score is the average of the individual feature scores.

Total score out of 100 is 1/4 * style score + 3/4 * code_score + bonus points. Bonus is typically in the range of 0-10 points.