Learn Git and GitHub without any code!
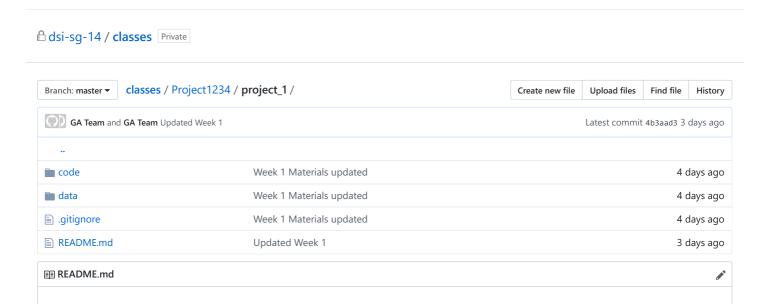
Using the Hello World guide, you'll start a branch, write comments, and open a pull request.

Read the guide

🔒 dsi-sg-14 / **classes** `Private`

| Branch: master ▾ | **classes** / **Project1234** / **project_1** / | | | Create new file | Upload files | Find file | History |

| 🔲 GA Team and GA Team Updated Week 1 | | Latest commit 4b3aad3 3 days ago |

..

📁 code	Week 1 Materials updated	4 days ago
📁 data	Week 1 Materials updated	4 days ago
📄 .gitignore	Week 1 Materials updated	4 days ago
📄 README.md	Updated Week 1	3 days ago

📖 README.md ✏️

🔴 Project 1: Standardized Testing, Statistical Summaries and Inference

Overview

Our first module in DSI covers:

- basic statistics (distributions, confidence intervals, hypothesis testing)
- many Python programming concepts
- programmatically interacting with files and directories
- visualizations
- EDA
- working with Jupyter notebooks for development and reporting

You might wonder if you're ready to start doing data science. While you still have **tons** to learn, there are many aspects of the data science process that you're ready to tackle. Project 1 aims to allow you to practice and demonstrate these skills.

For our first project, we're going to take a look at aggregate SAT and ACT scores and participation rates from each state in the United States. We'll seek to identify trends in the data and combine our data analysis with outside research to identify likely factors influencing participation rates and scores in various states.

Generally speaking, you will be asked to come up with a data science problem. Here's a specific prompt that should help you craft this statement:

> The new format for the SAT was released in March 2016. As an employee of the College Board - the organization that administers the SAT - you are a part of a team that tracks statewide participation and recommends where money is best spent to improve SAT participation rates. Your presentation and report should be geared toward **non-technical** executives with the College Board and you will use the provided data and outside research to make recommendations about how the College Board might work to increase the participation rate in a **state of your choice**.

Datasets

Provided Data

For this project, you'll have two provided datasets:

- 2017 SAT Scores
- 2017 ACT Scores

These data give average SAT and ACT scores by state, as well as participation rates, for the graduating class of 2017.

You can see the source for the SAT data here, and the source for the ACT data here. **Make sure you cross-reference your data with your data sources to eliminate any data collection or data entry issues.**

Additional Data

2018 state-by-state average results and participation for the SAT are available in PDF reports here. 2018 ACT state-by-state mean composite scores and participation rates are here .

This data has been compiled into CSV files which are also included in the *data* directory of this repo

Deliverables

All of your projects will comprise of a written technical report and a presentation. As we continue in the course, your technical report will grow in complexity, but for this initial project it will comprise:

- A Jupyter notebook that describes your data with visualizations & statistical analysis.
- A README markdown file the provides an introduction to and overview of your project.
- Your presentation slideshow rendered as a .pdf file. **NOTE**: Your entire Github repository will be evaluated as your technical report. Make sure that your files and directories are named appropriately, that all necessary files are included, and that no unnecessary or incomplete files are included.

For your first presentation, you'll be presenting to a **non-technical** audience. You should prepare a slideshow with appropriately scaled visuals to complement a compelling narrative. **Presentation duration will differ by market, so check with your local instructor.**

Technical Report Starter Code

Future projects will require you to decide on the entire structure of your technical report. Here, we provide you with starter code in a Jupyter notebook that will help to guide your data exploration and analysis. **If you choose to edit the core structure of this notebook, make sure you don't exclude any of the requested operations**.

Style Guide and Suggested Resources

Tim Dwyer (former DSI student and TA) put together this style guide. Some recommendations are geared toward future projects (which will include modeling and span multiple notebooks), but generally these are great recommendations.

Here's a link on how to give a good lightning talk, which provides some good recommendations for short presentations.

Here's a great summary of the main points of the book *Storytelling with Data*, which I can't recommend enough. Here's a blog post by the author about his guiding principles for visualizations.

Submission

Materials must be submitted by the beginning of class on April 20.

Your technical report will be hosted on Github Enterprise. Make sure it includes:

- A README.md (that isn't this file)
- Jupyter notebook(s) with your analysis (renamed to describe your project)

- Data files
- Presentation slides
- Any other necessary files (images, etc.)

Check with your local instructor for how they would like you to submit your repo for review.

Presentation Structure

- **Must be within time limit established by local instructor.**
- Use Google Slides or some other visual aid (Keynote, Powerpoint, etc).
- Consider the audience. Assume you are presenting to non-technical executives with the College Board (the organization that administers the SATs).
- Start with the **data science problem**.
- Use visuals that are appropriately scaled and interpretable.
- Talk about your procedure/methodology (high level, **CODE IS ALWAYS INAPPROPRIATE FOR A NON-TECHNICAL AUDIENCE**).
- Talk about your primary findings.
- Make sure you provide **clear recommendations** that follow logically from your analyses and narrative and answer your data science problem.

Be sure to rehearse and time your presentation before class.

Rubric

Your local instructor will evaluate your project (for the most part) using the following criteria. You should make sure that you consider and/or follow most if not all of the considerations/recommendations outlined below **while** working through your project.

Scores will be out of 21 points based on the 7 items in the rubric.
3 points per section

Score	Interpretation
0	*Project fails to meet the minimum requirements for this item.*
1	*Project meets the minimum requirements for this item, but falls significantly short of portfolio-ready expectations.*
2	*Project exceeds the minimum requirements for this item, but falls short of portfolio-ready expectations.*
3	*Project meets or exceeds portfolio-ready expectations; demonstrates a thorough understanding of every outlined consideration.*

Project Organization

- Are modules imported correctly (using appropriate aliases)?
- Are data imported/saved using relative paths?
- Does the README provide a good executive summary of the project?
- Is markdown formatting used appropriately to structure notebooks?
- Are there an appropriate amount of comments to support the code?
- Are files & directories organized correctly?
- Are there unnecessary files included?
- Do files and directories have well-structured, appropriate, consistent names?

Clarity of Message

- Is the problem statement clearly presented?
- Does a strong narrative run through the project?
- Does the student provide appropriate context to connect individual steps back to the overall project?
- Is it clear how the final recommendations were reached?
- Are the conclusions/recommendations clearly stated?

Python Syntax and Control Flow

- Is care taken to write human readable code?
- Is the code syntactically correct (no runtime errors)?
- Does the code generate desired results (logically correct)?
- Does the code follows general best practices and style guidelines?
- Are Pandas functions used appropriately?
- Does the student demonstrate mastery masking in Pandas?
- Does the student demonstrate mastery sorting in Pandas?

Data Cleaning and EDA

- Does the student fix data entry issues?
- Are data appropriately labeled?
- Are data appropriately typed?
- Are datasets combined correctly?
- Are appropriate summary statistics provided?
- Are steps taken during data cleaning and EDA framed appropriately?

Visualizations

- Are the requested visualizations provided?
- Do plots accurately demonstrate valid relationships?
- Are plots labeled properly?
- Plots interpreted appropriately?
- Are plots formatted and scaled appropriately for inclusion in a notebook-based technical report?

Research and Conceptual Understanding

- Were useful insights gathered from outside sources?
- Are sources clearly identified?
- Does the student provide appropriate interpretation with regards to descriptive and inferential statistics?

Presentation

- Is the problem statement clearly presented?
- Does a strong narrative run through the presentation building toward a final conclusion?
- Are the conclusions/recommendations clearly stated?
- Is the level of technicality appropriate for the intended audience?
- Is the student substantially over or under time?
- Does the student appropriately pace their presentation?
- Does the student deliver their message with clarity and volume?
- Are appropriate visualizations generated for the intended audience?
- Are visualizations necessary and useful for supporting conclusions/explaining findings?

In order to pass the project, students must earn a minimum score of 1 for each category.

- Earning below a 1 in one or more of the above categories would result in a failing project.
- While a minimum of 1 in each category is the required threshold for graduation, students should aim to earn at least an average of 1.5 across each category. An average score below 1.5, while it may be passing, means students may want to solicit specific feedback in order to significantly improve the project before showcasing it as part of a portfolio or the job search.

REMEMBER:

This is a learning environment and you are encouraged to try new things, even if they don't work out as well as you planned! While this rubric outlines what we look for in a *good* project, it is up to you to go above and beyond to create a *great* project. **Learn from your failures and you'll be prepared to succeed in the workforce**.