

# Part 1: HTML Bio

INTRODUCE YOURSELF! Using a Jupyter notebook with Markdown, create an HTML report that provides me with background information about you. Here is an [old example introduction to me](#) that you can base your outline on (disregard the homework problems at the bottom). You do not need to follow this outline exactly; however, this provides the fundamental output I am looking for. I do expect you to include an image so I know who you are! Be sure to use headings appropriately and feel free to incorporate features such as italics, bold font, lists, and HTML links. If you want to get saucy, show me you can write an equation, create a table, or embed a video; however, these are not required.

## What to include

The basics that I want you to include in your report are:

- Which one is you? (include a picture)
- A synopsis of who you are
- Academic Background
- Professional Background
- Experience with Python
- Experience with other analytic software

## Markdown references:

- [Jupyter Notebook docs](#)
- [Markdown docs](#)

## Submission

With your bio complete you will need to export the report as an HTML output (File >> Download as >> HTML) and you will upload the HTML file in the lab quiz.

# Part 2: Importing data

1. Import the blood\_transfusion.csv file (provided via Canvas).
  - Import the Pandas library with the alias `pd`.
  - What are the dimensions of this data (number of rows and columns)?
  - What are the data types of each column?
  - Are there any missing values?
  - Check out the first 10 rows? What are the `Class` values for the first 10 observations?
  - Check out the last 10 rows? What are the `Class` values for the last 10 observations?
  - Index for the 100th row and just the `Monetary` column. What is the value?
  - Index for just the `Monetary` column. What is the mean of this vector?
2. Let's dig into the data behind the story [Some People Are Too Superstitious To Have A Baby On Friday The 13th](#). You can download this data directly from the source by modifying the following

code with the URL -

[https://raw.githubusercontent.com/fivethirtyeight/data/master/births/US\\_births\\_2000-2014\\_SSA.csv](https://raw.githubusercontent.com/fivethirtyeight/data/master/births/US_births_2000-2014_SSA.csv)

```
url = "http://enter_url_here.html"
births = pd.read_csv(url)
```

This data contains U.S. births data for the years 2000 to 2014, as provided by the Social Security Administration. Use this data to answer the following questions:

- What are the dimensions of this data (number of rows and columns)?
  - Are there any missing values in this data?
  - Index for the 365th row. What is the date of this observation and what was the number of births in the US on this date?
  - Use indexing and the `.loc` accessor to get the first 31 days in our data frame for just the `births` column. Is the output a Series or a DataFrame. Now use the `.describe()` method on this output to compute various summary statistics such as min, max, mean, and median births across the first 31 days in the data.
3. Import the `PDI__Police_Data_Initiative__Crime_Incidents.csv` data (provided via Canvas). Data is taken from the [City of Cincinnati Open Data Portal website](https://data.cincinnati-oh.gov/safety/PDI-Police-Data-Initiative-Crime-Incidents/k59e-2pvf) (<https://data.cincinnati-oh.gov/safety/PDI-Police-Data-Initiative-Crime-Incidents/k59e-2pvf>), which you may need to read to place context in your answers.
- What are the dimensions of this data (number of rows and columns)?
  - Are there any missing values in this data? If so, how many missing values are in each column? Which column has the most missing values?
  - Index for the `DATE_REPORTED` column and apply the `.describe()` method to it. Which date has the most observations in this data set?
  - Index for the 100th observation in this data set. What was the date of this crime? What was the offense?