# CMSC6950

Computer Based Research Tools and Applications

git · · ·

million scales.

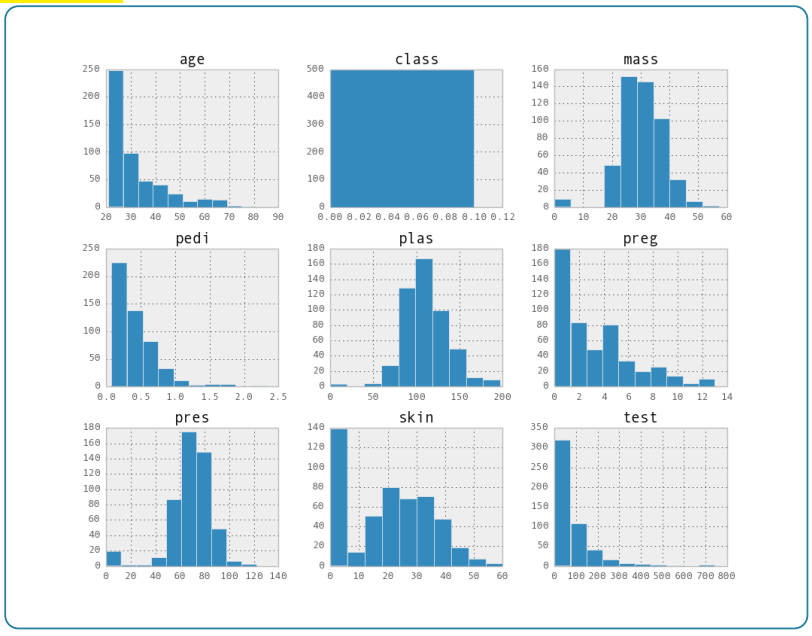
## Project specfication

- All projects are expected to demonstrate the following components:

1. Pick Your Project
    - Understand the project objectives
    - Goals
    - Tasks
    - Clarify roles.
    - Tasks divided fairly between the group team.
2. Download the Data
    - Fetching the data should be automated (create a function to do that).
    - Use the `requests` python library.
    - Use the `argparse` python library to write a command line program that takes arguments,

```
# This is just an example
$ download.py -month 2 -year 2010 -outfolder "./download" -outfile "file_2010_2.csv"
```

    - plot a histogram for each numerical attribute to understand the type of data you are dealing with. Example:




3. Data Cleaning

Create a few functions (a script `cleandata.py`) to take care of your raw data. The script should import the data from the data files (`.csv`) into more readable and useful information data.

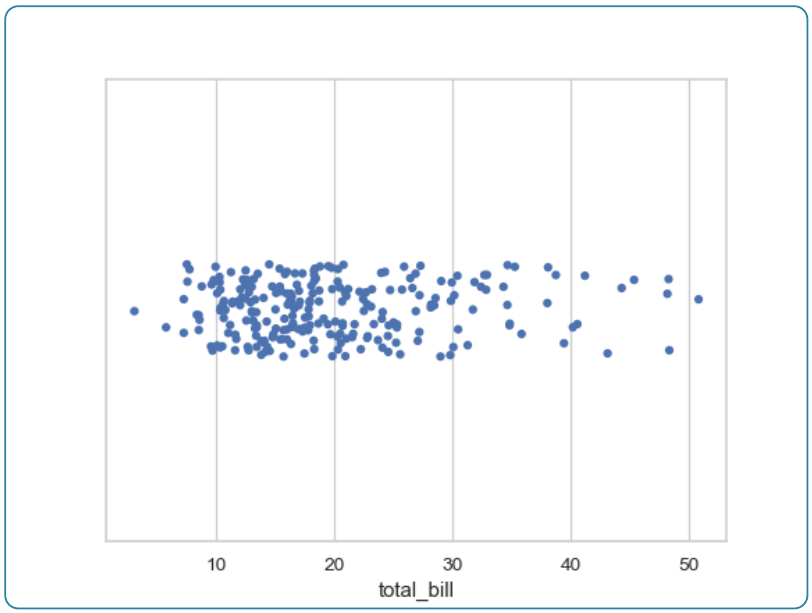
- Working with missing data (you could use `pandas.DataFrame.fillna`, `pandas.DataFrame.dropna`, and `pandas.DataFrame.drop` to accomplish these easily).
- Removing not needed columns in a DataFrame

# CMSC6950

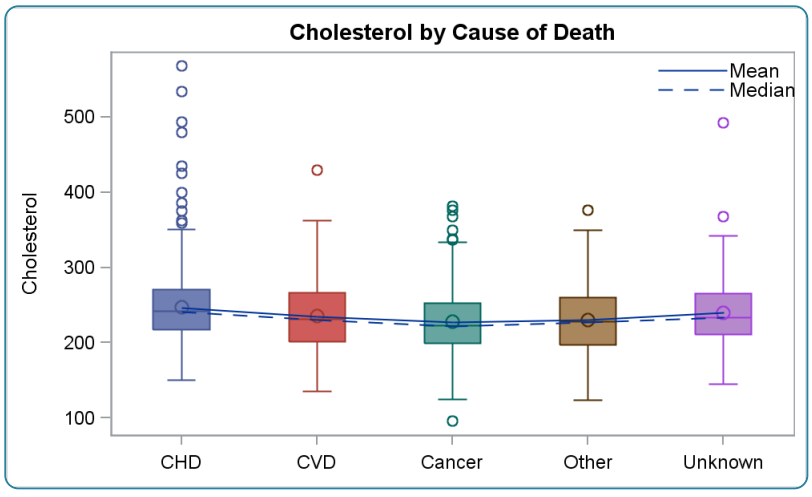
Computer Based Research Tools and Applications

The output from this script needs to be persistently stored (written to a file). Your choice on how to implement this storage.

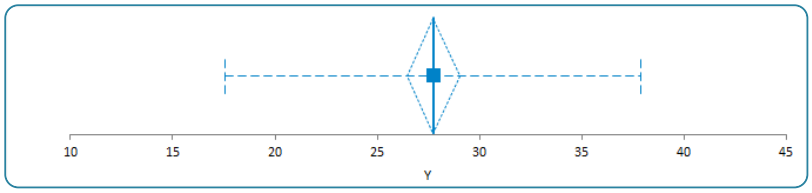
4. Analyze processing and plots
    - Visualize your data using two types of plots:
        1. Univariate plots: to understand each attribute. (**two plots**)
            - Dot plot (a strip plot)




            - Box plot



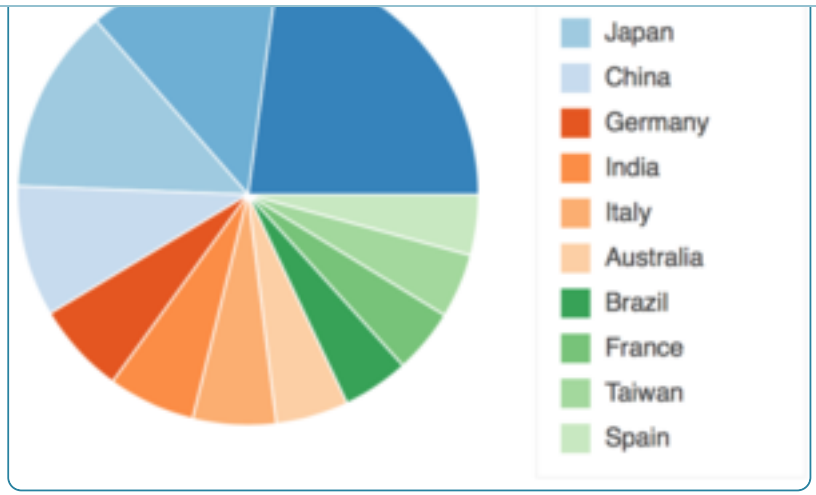

            - Mean plot: shows the mean and standard deviation of the data.

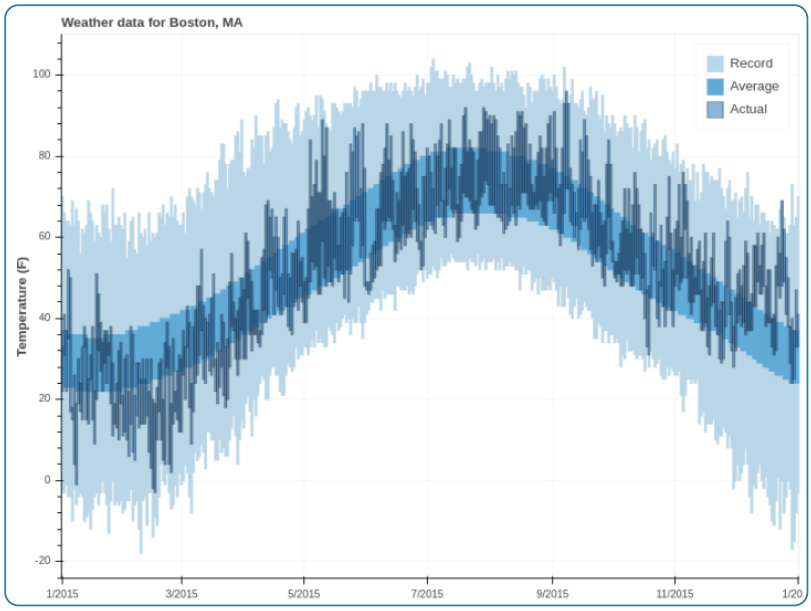



        2. Multivariate plots: to understand the relationships between attributes. (**two plots**)

# CMSC6950

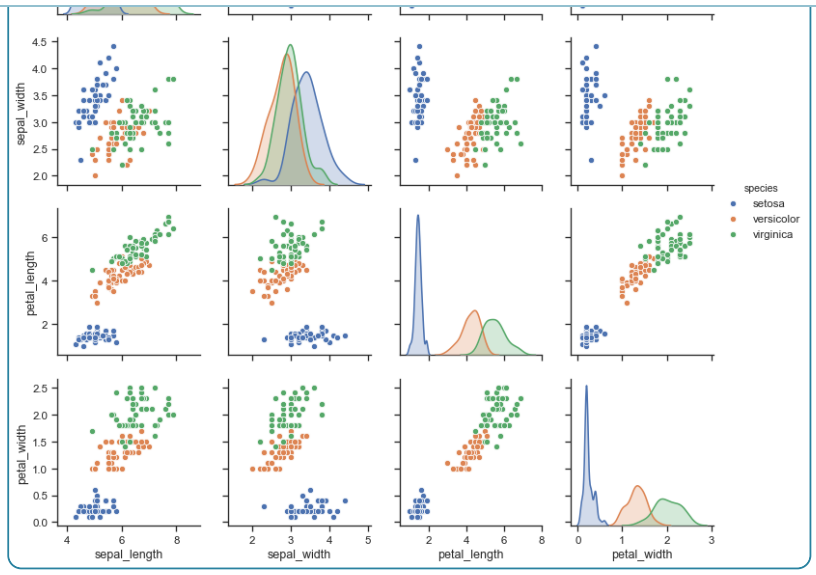
Computer Based Research Tools and Applications

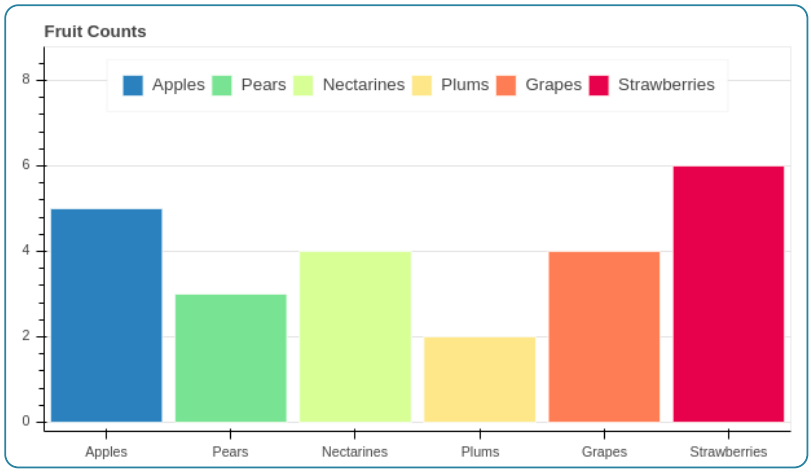


# CMSC6950

Computer Based Research Tools and Applications

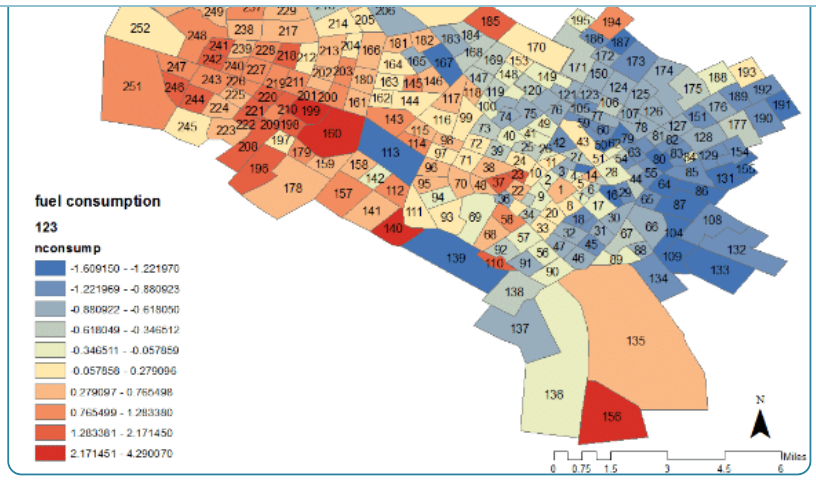


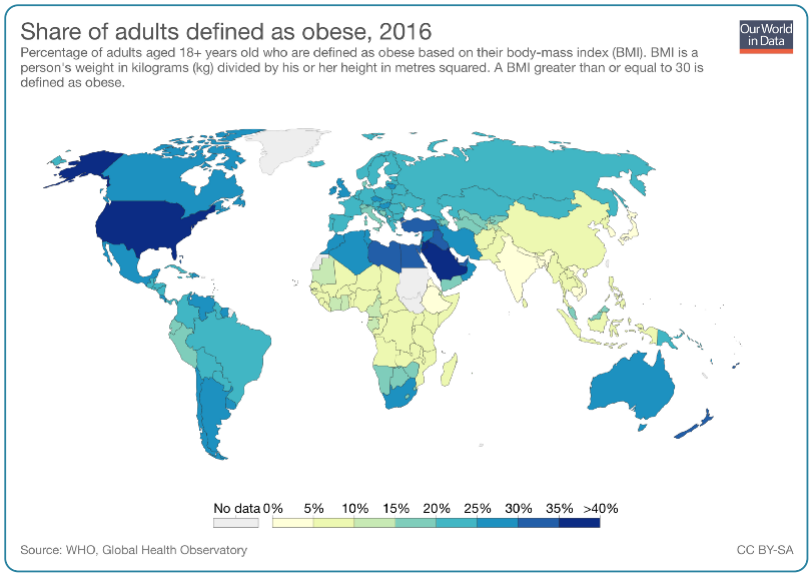




- Create standalone bokeh plots embedded in your HTML presentation so that users can interactively select with the data (hover points).
- ~~Plot on maps. (**single plot**)~~ (NOT REQUIRED)
  Example:

# CMSC6950

Computer Based Research Tools and Applications




5. Use version control (git) and collaboration tools (GitLab) throughout this project. **Important:** make sure you have properly set up your git configuration with your MUN email address. We will use this information to assess your individual contributations to this team project. Team members should all collaborate on a single gitlab repository. The use of branches is permitted but do not use gitlab forks and pull requests.

6. You should automatically creates and manages a virtualenv for your project (**Pipenv**). Specify versions of a package,

```
pipenv install flask==0.12.1
```

~~7. Create a LaTeX report summarizing the results of your project.~~

8. Create a web based presenation for your results. The remark-js library is nice for doing HTML based presenations Host your presenation on GitLab or GitHub pages.

9. Implement your entire workflow as a **Makefile**. Ensure that your entire project is reproducible.

# CMSC6950

Computer Based Research Tools and Applications

git · · ·

**Readme.md** file to explain how to use/build your project. See chapter 19 for some ideas on documentation. While the use of **docstrings** in your code is encouraged, you do not need to set up an automated documenation framework like Sphinx.

## Evaluation

Assessment will based on both the actual results and demonstration of good scientific computing techniques as discussed within the course and the textbook.

- Minimum tasks 65%
- Individual/team participation 20% (based on git contributions)
- Presentation 15%

# Previous Projects

1. Growing Degree Days