

## Repository Information

### Where to Download Data

Data from SHAPE America was transferred from a PDF into a CSV which was then coded into a pandas DataFrame. For ease, the SHAPE America CSV file is provided in the repository along with the original PDF file from which the data was obtained. The CDC data was downloaded from the CDC website and made into a CSV file. That data is downloadable at:

<https://nccd.cdc.gov/youthonline/App/Results.aspx?TT=B&OUT=0&SID=HS&QID=QNPA0DAY&LID=LL&YID=RY&LID2=&YID2=&COL=&ROW1=&ROW2=&HT=&LCT=&FS=&FR=&FG=&FA=&FI=&FP=&FSL=&FRL=&FGL=&FAL=&FIL=&FPL=&PV=&TST=&C1=&C2=&QP=&DP=&VA=CI&CS=Y&SYID=&EYID=&SC=&SO=>. Again, for ease of use the CSV file is provided in this repository. Finally, TIGER/Line shape files for the US States were downloaded from the US Census page. This can be done by going to <https://www.census.gov/cgi-bin/geo/shapefiles/index.php> and setting the year to 2020 and the layer type to State. The output file should be called “tl\_2020\_us\_state.zip”.

### Repository Script Sequence, Inputs & Outputs

1. state\_reqs.py
  - a. **Inputs:** SHAPE America info (“P1.csv”, “P2.csv”, and “P3.csv”)
  - b. **Outputs:** High school policy data with scores (“hs\_reqs.csv”)
2. analyze\_reqs.py
  - a. **Inputs:** “hs\_reqs.csv”
  - b. **Outputs:** There is no output for this script. It was used to see which, if any, policies were highly related.
3. state\_health.py
  - a. **Inputs:** CDC Health Info (“YRBS2019\_HS.csv”)
  - b. **Outputs:** High school health info by state (“state\_health.csv”) and a pair plot (“scatterhealth.png”) that looked to see if any health behaviors are highly related.
4. Analyze\_health.py
  - a. **Inputs:** “state\_health.csv”, “hs\_reqs.csv”, a CSV files of labels that reduces overlap when labelling states in scatter plots (“labels.csv”)
  - b. **Outputs:** multi-panel heatmap with health characteristics by state (“HealthbyState.png”), health and scores merged into one file (“health\_data.csv”), and 6 scatter plots with quadrants (“scores\_obesity.png”, “scores\_not1.png”, “scores\_suicide.png”, “scores\_1to5.png”, “score\_sleep.png”, “scores\_5over.png”)
5. state\_data.py
  - a. **Inputs:** “health\_data.csv”, “hs\_reqs.csv”, state shape files downloaded from the Census (see above for instructions): “tl\_2020\_us\_state.zip”
  - b. **Outputs:** CSV file with all health data and state policy information (“state\_data.csv”), the same file formatted for Stata (“state\_data.dta”), and the same file formatted for use in GIS (“state\_data.gpkg”).
6. corrs.py

- a. **Inputs:** "correlations.csv", derived using Stata.
- b. **Outputs:** A heat map that visually displays the correlations between requirements and health outcomes.